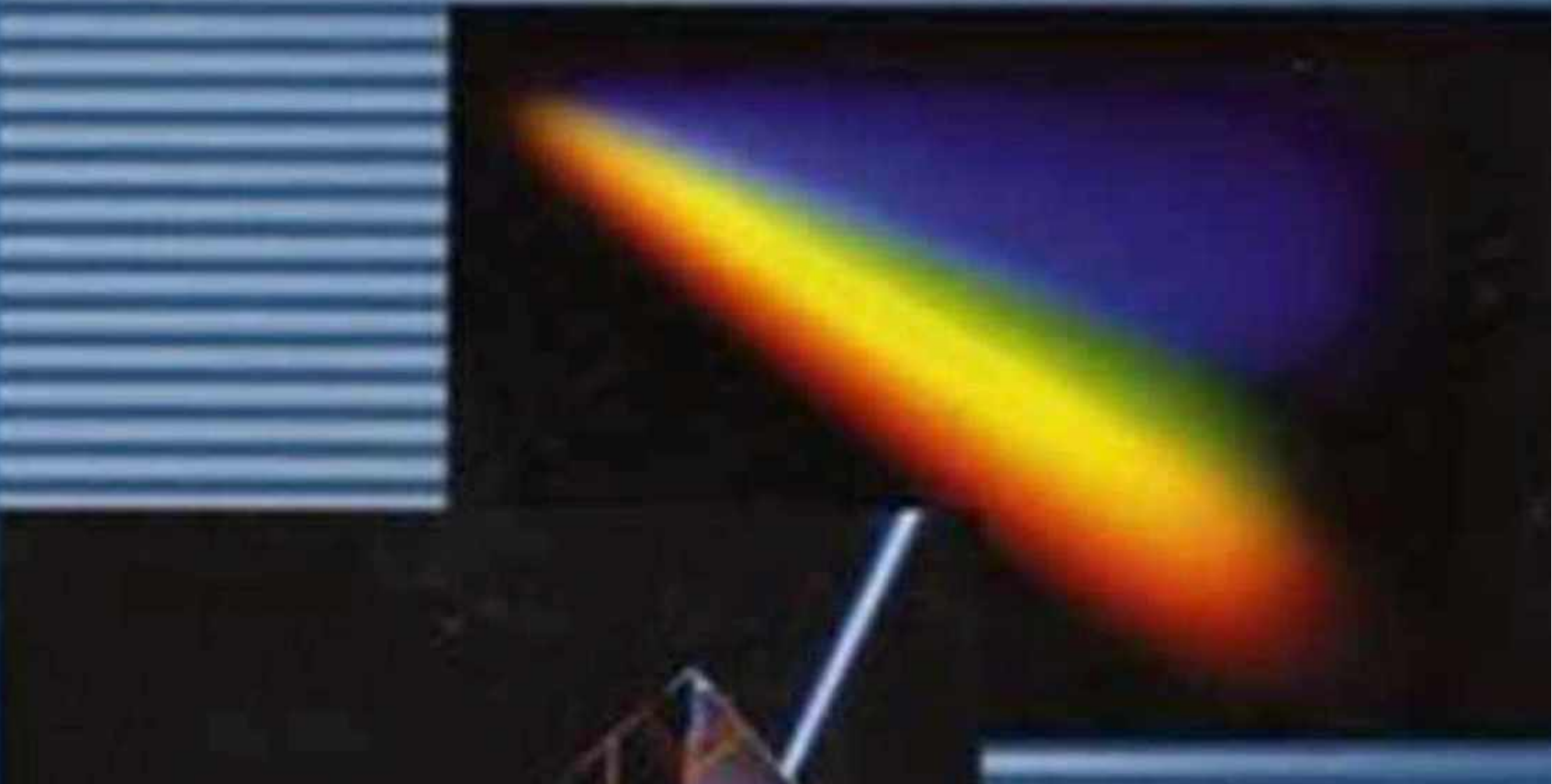
ОБЩИЙ КУРС ФИЗИКИ

Д.В. Сивухин

ОПТИКА

ТОМ IV



УДК 535
ББК 22.34
С34

Сивухин Д. В. **Общий курс физики.** Учеб. пособие: Для вузов. В 5 т. Т. IV. **Оптика.** — 3-е изд., стереот. — М.: ФИЗМАТЛИТ, 2005. — 792 с. — ISBN 5-9221-0228-1.

Четвертый том курса физики, широко известного у нас в стране и за рубежом. Книга написана на основе лекций, которые в течение ряда лет читались автором студентам Московского физико-технического института. Основное внимание уделено выяснению физического смысла и содержания основных законов и понятий оптики, установлению границ применимости этих законов, развитию у студентов навыков физического мышления и умения ставить и решать конкретные задачи.

Второе издание четвертого тома вышло в 1985 г.

Для студентов физических и математических факультетов университетов, физико-технических и инженерно-физических институтов, а также вузов, где физика является основной дисциплиной.

ISBN 5-9221-0228-1 (Т. IV)
ISBN 5-9221-0229-X

© ФИЗМАТЛИТ, 1985, 2002, 2005

ОГЛАВЛЕНИЕ

Предисловие 7

ГЛАВА I

ВВЕДЕНИЕ

§ 1. Предмет оптики 9
§ 2. Геометрическая оптика 12
§ 3. Эволюция представлений о природе света 20
§ 4. Искривление световых лучей в неоднородных сферах 33
§ 5. Плоские электромагнитные волны 37
§ 6. Предельный переход от волновой оптики к геометрической . . . 44
§ 7. Принцип Ферма 49
§ 8. Групповая скорость 57

ГЛАВА II

ГЕОМЕТРИЧЕСКАЯ ТЕОРИЯ ОПТИЧЕСКИХ ИЗОБРАЖЕНИЙ

§ 9. Понятие оптического изображения 67
§ 10. Преломление на сферической поверхности. Сферические зеркала и тонкие линзы 73
§ 11. Общие свойства центрированных оптических систем 78
§ 12. Сложение центрированных систем. Толстые линзы 89
§ 13. Ограничение лучей при помощи диафрагм 96
§ 14. Астигматические пучки лучей. Каустика 101
§ 15. Геометрические аберрации центрированных систем 106
§ 16. Хроматическая аберрация 113
§ 17. Условие отсутствия дисторсии 120
§ 18. Условие синусов Аббе 122
§ 19. Теорема косинусов. Стигматические изображения широкими пучками лучей 129
§ 20. Об абсолютно оптических инструментах 135
§ 21. Глаз и зрение 139
§ 22. Фотометрические понятия и единицы 152
§ 23. Яркость и освещенность оптического изображения. Нормальное увеличение 163
§ 24. Оптические инструменты 172
§ 25. Электрические и магнитные линзы 190

ГЛАВА III
ИНТЕРФЕРЕНЦИЯ СВЕТА

§ 26. Общие сведения об интерференции 199
§ 27. Классические интерференционные опыты 210
§ 28. Влияние размеров источника света. Пространственная когерентность 217
§ 29. Спектральное разложение 225
§ 30. Влияние немонохроматичности света 230
§ 31. Корреляция и когерентность света 234
§ 32. Теорема Ван-Циттера–Цернике 240
§ 33. Интерференция в пленках и пластинках 242
§ 34. Интерферометр Жамена 249
§ 35. Интерферометр Майкельсона 256
§ 36. Многолучевая интерференция 258
§ 37. Стоячие световые волны 267
§ 38. Излучение Вавилова–Черенкова 271

ГЛАВА IV
ДИФРАКЦИЯ СВЕТА

§ 39. Принцип Гюйгенса–Френеля. Зоны Френеля 277
§ 40. Дифракция на оси от круглого отверстия и экрана. Зонная пластинка 284
§ 41. Метод Френеля решения дифракционных задач. Дифракция Фраунгофера и Френеля 291
§ 42. Зоны Шустера и спираль Корню 298
§ 43. Принцип Гюйгенса в формулировке Кирхгофа 304
§ 44. Дифракция Фраунгофера на щели 309
§ 45. Дифракция Фраунгофера на отверстиях 315
§ 46. Дифракционная решетка 320
§ 47. Дифракционная решетка как спектральный прибор 329
§ 48. Эшелон Майкельсона и интерференционные спектральные приборы 334
§ 49. Разрешающая способность призмы 339
§ 50. Действие спектрального аппарата на световые импульсы 344
§ 51. Вогнутая отражательная решетка 350
§ 52. Дифракция на решетке как краевая задача 353
§ 53. Примеры на применение метода Рэлея 359
§ 54. Голография 363
§ 55. Световое поле вблизи фокуса 374
§ 56. Разрешающая способность телескопа и микроскопа 378
§ 57. Теория и демонстрационные опыты Аббе 388
§ 58. Телескоп без объектива. Получение изображений с помощью малых отверстий 395
§ 59. Фазовый контраст 399
§ 60. Измерение угловых диаметров звезд 402
§ 61. Дифакция на двумерных и трехмерных решетках. Дифракция рентгеновских лучей 407

ГЛАВА V

ОТРАЖЕНИЕ И ПРЕЛОМЛЕНИЕ СВЕТА

§ 62. Поляризованный и естественный свет 420
§ 63. Число независимых граничных условий в электромагнитной теории света 423
§ 64. Геометрические законы отражения и преломления волн 426
§ 65. Формулы Френеля 430
§ 66. Полное отражение 436
§ 67. Прохождение света через плоскопараллельную пластинку. Просветление оптики 441
§ 68. Распространение света в среде с точки зрения молекулярной оптики 450
§ 69. Вывод формул Френеля в молекулярной оптике 455
§ 70. Отступления от формул Френеля 459

ГЛАВА VI

ОПТИКА МЕТАЛЛОВ

§ 71. Уравнения Максвелла и волны в металлах 466
§ 72. Геометрические законы отражения и преломления света на границе металла 469
§ 73. Формулы Френеля. Измерение оптических констант металлов . 472
§ 74. Аномальный скин-эффект и эффективная диэлектрическая проницаемость 478

ГЛАВА VII

КРИСТАЛЛООПТИКА

§ 75. Плоские волны в кристаллах 481
§ 76. Оптически одноосные кристаллы 483
§ 77. Поляризационные устройства 491
§ 78. Анализ поляризованного света 499
§ 79. Интерференция поляризованных лучей 508
§ 80. Нормальные скорости и поляризация волн в двуосных кристаллах 520
§ 81. Лучи, волновые нормали и связь между ними 528
§ 82. Коническая рефракция 537
§ 83. Замечания об отражении и преломлении света на границе кристаллов 543

ГЛАВА VIII

МОЛЕКУЛЯРНАЯ ОПТИКА

§ 84. Классическая теория дисперсии света 548
§ 85. Понятие о дисперсионной формуле квантовой механики 559
§ 86. Методы экспериментального исследования аномальной дисперсии 564
§ 87. Дисперсия плазмы 569
§ 88. Средняя плотность электромагнитной энергии в диспергирующих средах 573
§ 89. Поглощение света и уширение спектральных линий 576
§ 90. Двойное преломление света в электрическом и магнитном полях 583
§ 91. Линейный электрооптический эффект Поккельса 594

§ 92. Эффект Зеемана . . . 596
§ 93. Понятие об эффекте Штарка . . . 602
§ 94. Вращение плоскости поляризации . . . 604
§ 95. Магнитное вращение плоскости поляризации . . . 611
§ 96. Временная и пространственная дисперсия. Теория естественной оптической активности . . . 617
§ 97. О тепловых флуктуациях . . . 626
§ 98. Рассеяние света . . . 632
§ 99. Явление Мандельштама–Бриллюэна . . . 643
§ 100. Комбинационное рассеяние света . . . 649

ГЛАВА IX

ТЕОРИЯ ОТНОСИТЕЛЬНОСТИ

§ 101. Принцип относительности в ньютоновской и релятивистской механике . . . 655
§ 102. Опыт Майкельсона . . . 659
§ 103. Независимость скорости света от движения источника . . . 665
§ 104. Понятие одновременности . . . 667
§ 105. Преобразование координат и времени в теории относительности 671
§ 106. Лорентцово сокращение длины и замедление времени . . . 680
§ 107. Эффект Допплера и аберрация света . . . 688
§ 108. Эффект Допплера в акустике и теории эфира . . . 695
§ 109. Замедление хода часов в гравитационном поле . . . 699
§ 110. Сложение скоростей в теории относительности . . . 702
§ 111. Релятивистская механика . . . 706

ГЛАВА X

ТЕПЛОВОЕ ИЗЛУЧЕНИЕ

§ 112. Равновесное излучение в полости . . . 713
§ 113. Закон Кирхгофа . . . 717
§ 114. Формула Кирхгофа–Клаузиуса . . . 721
§ 115. Закон Стефана–Больцмана . . . 723
§ 116. Теорема и закон смещения Вина . . . 726
§ 117. Формула Рэлея–Джинса . . . 731
§ 118. Формула Планка . . . 737
§ 119. Спонтанное и индуцированное излучение . . . 743

ГЛАВА XI

ЛАЗЕРЫ И НЕЛИНЕЙНАЯ ОПТИКА

§ 120. Принципы работы лазера . . . 748
§ 121. Рубиновый лазер . . . 754
§ 122. Гелий-неоновый лазер . . . 762
§ 123. Нелинейная поляризация среды . . . 765
§ 124. Первое приближение. Оптическое детектирование. Генерация вторых гармоник, суммарной и разностной частот . . . 768
§ 125. Второе приближение. Самофокусировка . . . 773
§ 126. Параметрическая генерация света . . . 776

Именной указатель . . . 780
Предметный указатель . . . 783

ПРЕДИСЛОВИЕ

Этот — четвертый — том общего курса физики посвящен физической оптике и является естественным продолжением предыдущего тома, в котором излагается учение об электрических и магнитных явлениях. Физическая оптика рассматривается в нем преимущественно с волновой (конечно, электромагнитной) точки зрения. Вопросы квантовой оптики затрагиваются лишь частично. Дается представление о фотонах и процессе излучения как о квантовом переходе атомных систем из одного энергетического состояния в другое. Это необходимо для введения понятия индуцированного излучения и объяснения принципов работы лазеров. Однако систематическое изложение основ квантовой оптики, в той мере, в какой это возможно сделать в рамках общей физики, а также относящихся сюда квантовых явлений (фотоэффект, эффект Комптона, спектральные закономерности, люминесценция, эффект Зеемана, эффект Штарка и пр.), предполагается дать в пятом томе, где будет излагаться атомная физика в широком смысле этого слова.

Вопросы геометрической оптики собраны в первых двух главах курса, чтобы в дальнейшем можно было ссылаться на них при изложении интерференции, дифракции и других разделов физической оптики. Геометрическая оптика излагается не как математическая, а как физическая дисциплина — как приближенный предельный случай волновой оптики. Тем самым четко определяются границы ее применимости. С целью простоты в основу обоснования геометрической оптики положено скалярное волновое уравнение. Хотя в общем случае неоднородной среды оно и неверно, но даже в этом случае при рассмотрении предельного перехода к геометрической оптике оно приводит к правильным результатам. Конечно, на основе скалярного уравнения ничего нельзя сказать относительно вращения плоскости поляризации луча в неоднородной среде. Для этого надо было бы положить в основу векторные уравнения Максвелла. Но это, ничего не меняя в идейном отношении, потребовало бы довольно громоздких вычислений. Существенно, что скалярное волновое уравнение правильно передает основные закономерности распространения волн не только в однородных, но и в неоднородных средах. Геометрическая же оптика получается из него в предельном случае коротких волн, длины которых пренебрежимо малы по сравнению с характерными размерами, определяющими распространение света в среде.

Впрочем, обоснование геометрической оптики при всей его важности может быть опущено при первом чтении книги.

После изложения основ геометрической оптики во второй главе излагается геометрическая теория оптических изображений – главным образом с принципиальной стороны. Большая часть материала этой главы при первом чтении может быть также опущена. Достаточно ограничиться минимумом: получением оптических изображений в параксиальном приближении и основными понятиями фотометрии. К остальным вопросам можно обращаться по мере надобности в процессе изучения физической оптики.

Основное содержание курса составляет физическая оптика, изложение которой начинается с главы III. В нее входят также специальная теория относительности, краткое изложение принципов работы оптических квантовых генераторов (лазеров) и элементов нелинейной оптики. Как и в предыдущих томах курса, главное внимание здесь обращено на выяснение физического смысла и содержания оптических явлений, на связь между ними и с общими принципами физики. Автор стремился к идейной простоте изложения, но старался избегать вульгаризации.

Вопросы истории, экспериментальное обоснование законов оптики, применения ее в технике и других науках затронуты лишь постольку, поскольку это необходимо для уяснения основных явлений и принципов оптики. Из-за недостатка места даже описанию демонстраций в этом томе уделено меньше внимания, чем в предыдущих томах курса. Впрочем, демонстрации, осуществленные в Московском физико-техническом институте моими лекционными ассистентами Е.Н. Морозовым, М.И. Маклаковым, В.П. Молчановым, В.А. Кузнецовой, которым я глубоко благодарен, сыграли немалую роль при написании и этого тома курса физики.

Рукопись этого тома была тщательно просмотрена рецензентами — профессором С.С. Герштейном, профессором И.С. Горбанем и сотрудниками руководимой им кафедры экспериментальной физики Киевского государственного университета им. Т.Г. Шевченко, а также доктором физико-математических наук Г.П. Пекой. Их справедливые критические замечания учтены при окончательном редактировании книги. Всем этим лицам автор приносит глубокую благодарность.

Д.В. Сивухин

ГЛАВА I

ВВЕДЕНИЕ

---

§ 1. Предмет оптики

Оптика, точнее — *физическая оптика*, есть раздел физики, изучающий свойства и физическую природу света, а также его взаимодействие с веществом. Под светом понимают не только *видимый свет*, но и примыкающие к нему широкие области спектра электромагнитного излучения — *инфракрасную* и *ультрафиолетовую*. Различные участки спектра электромагнитного излучения отличаются друг от друга *длиной волны* λ и *частотой* ν — величинами, характеризующими не только *волновые*, но и *квантовые свойства* электромагнитного излучения. Электромагнитный спектр принято делить на *радиоволны*, *инфракрасное*, *видимое*, *ультрафиолетовое*, *рентгеновское* и *гамма-излучения*. Эти участки спектра различаются не по своей физической природе, а *по способу генерации и приема излучения*. Поэтому между ними нет резких переходов, сами участки перекрываются, а границы между ними условны.

Радиоволнами называются электромагнитные излучения, длины волн которых превосходят примерно 0,1 мм. Их принято делить на: 1) *сверхдлинные волны* с длиной волны $\lambda > 10$ км (частота $\nu < 30$ кГц); 2) *длинные волны* ($\lambda = 10 - 1$ км, $\nu = 30 - 300$ кГц); 3) *средние волны* ($\lambda = 1$ км $- 100$ м, $\nu = 300$ кГц $-$ 3 МГц); 4) *короткие волны* ($\lambda = 100 - 10$ м, $\nu = 3 - 30$ МГц); 5) *ультракороткие волны* ($\lambda < 10$ м, $\nu > 30$ МГц). Последние в свою очередь принято подразделять на *метровые*, *дециметровые*, *миллиметровые* и *субмиллиметровые* или *микрометровые*. Волны с длиной $\lambda < 1$ м ($\nu > 300$ МГц) принято также называть *микроволнами* или *волнами сверхвысоких частот* (СВЧ). Из-за больших значений λ распространение радиоволн можно рассматривать *феноменологически* без учета атомистического строения среды. Исключение составляют только самые короткие радиоволны, примыкающие к инфракрасному участку спектра. Практически не сказываются и квантовые свойства радиоизлучения.

Видимое, инфракрасное и ультрафиолетовое излучения составляют так называемую *оптическую область спектра* в широком смысле этого слова. Выделение такой области обусловлено не только близостью соответствующих участков спектра, но и сходством методов и приборов, применяющихся для ее исследования и разработанных исторически главным образом при изучении видимого света (линзы и зеркала для фокусировки излучения, призмы, дифракционные решетки, интерференционные приборы для исследования спектрального состава излучения и пр.). *Оптический спектр* занимает диапазон от условной границы инфракрасного излучения ($\lambda = 2$ мм, $\nu = 1{,}5 \cdot 10^{11}$ Гц) до условной коротковолновой границы ультрафиолета ($\lambda = 10^{-6}$см $= 10$ нм, $\nu = 3 \cdot 10^{16}$ Гц), что составляет примерно 18 октав[1]. Видимое излучение занимает приблизительно одну октаву ($\lambda = 400$ – 760 нм), ультрафиолет — 5 октав ($\lambda = 10$ – 400 нм), инфракрасное излучение — 11 октав ($\lambda = 760$ нм – 2 мм). В оптической области спектра частоты ν уже перестают быть малыми по сравнению с собственными частотами атомов и молекул, а длины волн большими по сравнению с молекулярными размерами и межмолекулярными расстояниями. Благодаря этому в этой области становятся существенными явления, обусловленные атомистическим строением вещества. По той же причине, наряду с волновыми, проявляются и квантовые свойства света. Энергия светового кванта определяется выражением

$$\mathcal{E} = h\nu, \tag{1.1}$$

где $h = 6{,}63 \cdot 10^{-27}$ эрг·с — *постоянная Планка*. Полезно заметить, что для длины волны $\lambda = 1000$ нм энергия соответствующего кванта составляет $\mathcal{E} = 1{,}23$ эВ, или приблизительно один электрон-вольт. На концах видимого спектра ($\lambda_{\text{кр}} = 760$ нм, $\lambda_{\text{фл}} = 400$ нм) для энергии кванта формула (1.1) дает $\mathcal{E}_{\text{кр}} \approx 1{,}6$ эВ, $\mathcal{E}_{\text{фл}} \approx 3$ эВ.

В области рентгеновского и гамма-излучения на первый план выступают *квантовые свойства излучения*. Рентгеновское излучение возникает при торможении быстрых заряженных частиц (электронов, протонов и пр.), а также в результате процессов, происходящих внутри электронных оболочек атомов. Гамма-излучение появляется в результате процессов, происходящих внутри атомных ядер, а также в результате превращения элементарных частиц. Оно появляется и при торможении быстрых заряженных частиц. Характерные энергии гамма-квантов — порядка одного или нескольких МэВ. Границы областей рентгеновского и гамма-излучения, в особенности коротковолновые, могут быть определены лишь весьма условно. Для общей ориентировки можно принять, что энергия рентгеновских квантов лежит в пределах 20 эВ – 1 МэВ ($\lambda = 50 - 10^{-3}$ нм), а энергия гамма-квантов — больше 0,1 МэВ ($\lambda < 10^{-2}$ нм).

---

[1] Октавой называется интервал частот между произвольной частотой ω и ее гармоникой 2ω.

Следует заметить, что волновые и квантовые закономерности являются общими для всего спектра электромагнитного излучения. Только, в зависимости от длины волны, на первый план выступают разные явления, разные методы исследования и разные практические применения. Поэтому на оптику нельзя смотреть как на замкнутую дисциплину, изучающую только оптическую область спектра, отделенную от других областей резкими границами. Закономерности и результаты, найденные в этих и других областях, могут оказаться применимыми в оптической области спектра и обратно.

Практическое значение оптики и ее влияние на другие отрасли знания исключительно велики. Изобретение телескопа и спектроскопа открыло перед человеком удивительнейший и богатейший мир явлений, происходящих в необъятной Вселенной. Изобретение микроскопа произвело революцию в биологии. Фотография помогла и продолжает помогать чуть ли не всем отраслям науки. Одним из важнейших элементов научной аппаратуры является линза. Без нее не было бы микроскопа, телескопа, спектроскопа, фотоаппарата, кино, телевидения и т.п. Не было бы очков, и многие люди, которым перевалило за пятьдесят лет, были бы лишены возможности читать и выполнять многие работы, связанные со зрением.

Область явлений, изучаемая физической оптикой, весьма обширна. Оптические явления теснейшим образом связаны с явлениями, изучаемыми в других разделах физики, а оптические методы исследования относятся к наиболее тонким и точным. Поэтому неудивительно, что оптике на протяжении длительного времени принадлежала ведущая роль в очень многих фундаментальных исследованиях и развитии основных физических воззрений. Достаточно сказать, что обе основные физические теории текущего столетия — теория относительности и теория квантов — зародились и в значительной степени развились на почве оптических исследований. Изобретение лазеров открыло новые широчайшие возможности не только в оптике, но и ее приложениях в различных отраслях науки и техники.

§ 2. Геометрическая оптика

1. Простейшие оптические явления, например возникновение теней и получение изображений в оптических приборах, могут быть поняты в рамках так называемой *геометрической оптики.* В основу формального построения последней можно положить четыре закона, установленных опытным путем: 1) *закон прямолинейного распространения света*; 2) *закон независимости световых пучков*; 3) *закон отражения* и 4) *закон преломления света.* Для понимания более сложных явлений нужна уже *физическая оптика*, рассматривающая эти явления в связи с *физической природой света.* Физическая оптика позволяет, в частности, не только вывести все законы геометрической оптики, но и установить *границы их применимости.* Без знания этих границ формальное применение законов геометрической оптики может в конкретных случаях привести к результатам, противоречащим наблюдаемым явлениям. Поэтому нельзя ограничиться чисто формальным построением геометрической оптики, а необходимо смотреть на нее как на *раздел физической оптики.* Обоснование геометрической оптики будет дано постепенно на протяжении нашего курса. Сейчас же ограничимся формулировкой перечисленных четырех опытных законов.

2. Согласно закону прямолинейного распространения, *свет в прозрачной однородной среде распространяется по прямым линиям.* Опытным доказательством этого закона могут служить резкие тени, отбрасываемые непрозрачными телами, освещаемыми точечными источниками света, т.е. источниками, размеры которых весьма малы по сравнению с размерами освещаемого тела и расстоянием до него. Непрозрачный предмет AB (рис. 1), поставленный на пути светового пучка от точечного источника S, не пропускает свет в пространство за этим предметом, ограниченное боковой поверхностью конуса SCD, касающейся краев предмета AB. Однако на распределение света вне этого

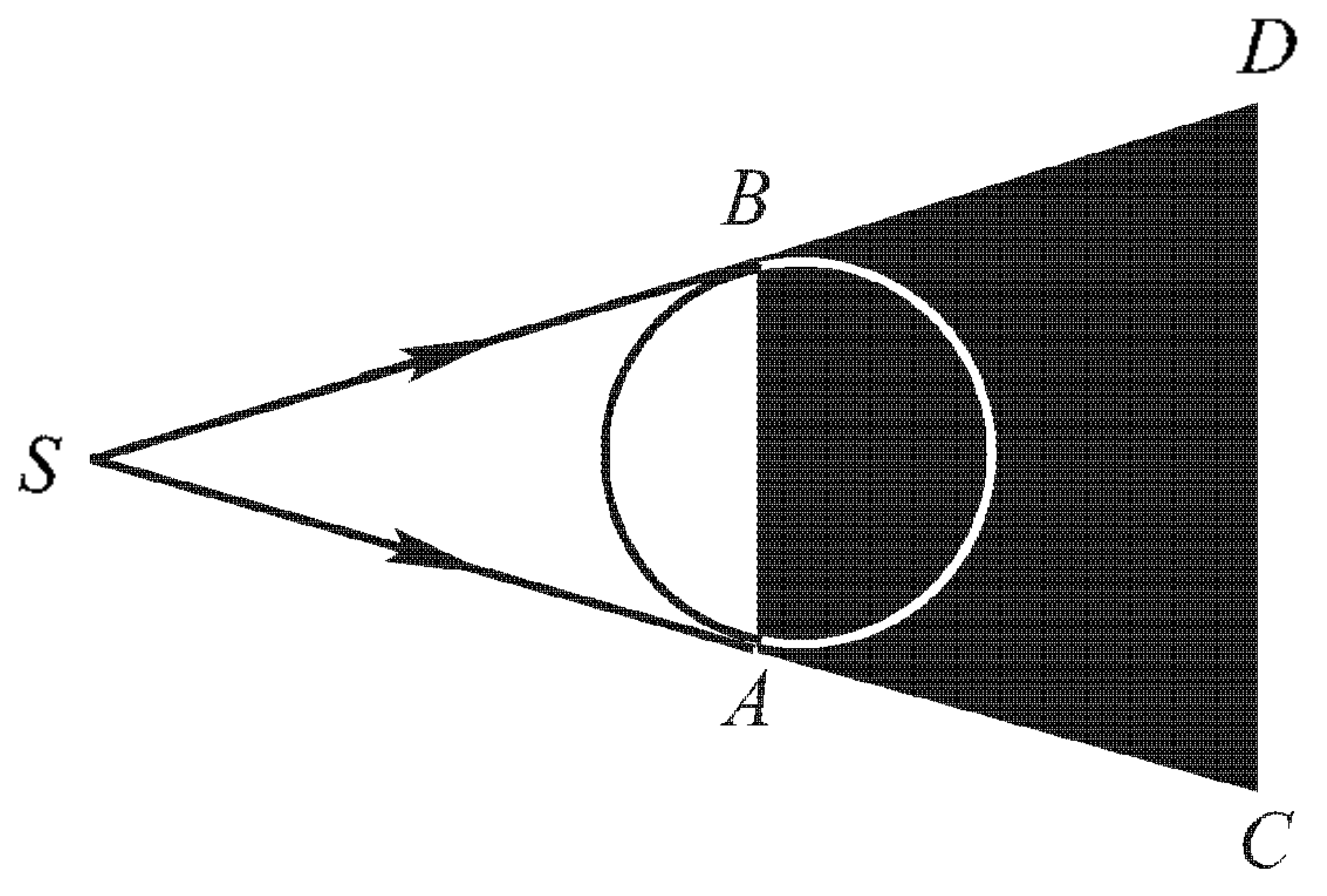


Рис. 1

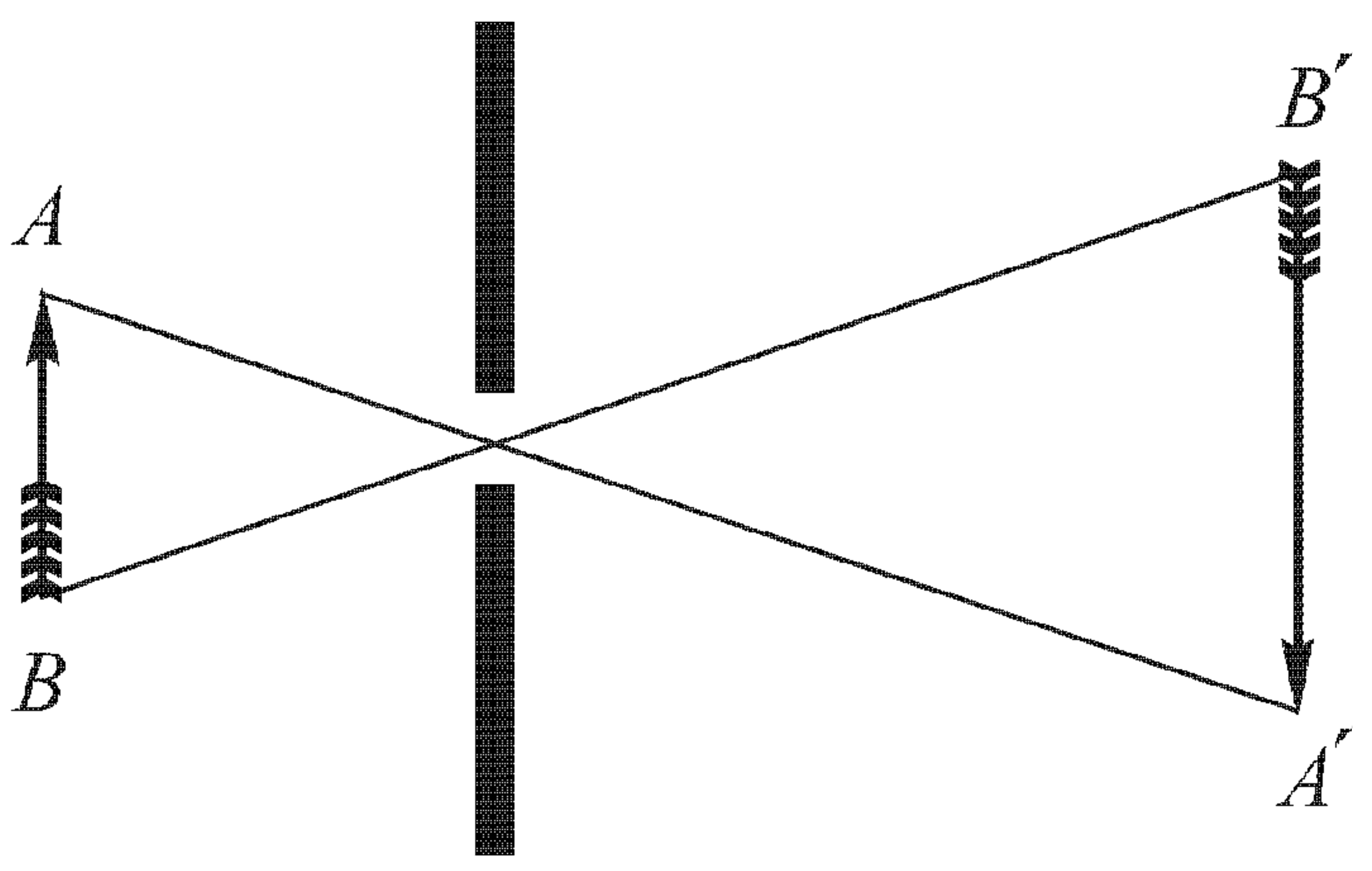


Рис. 2

пространства присутствие предмета AB не оказывает никакого влияния. Это и значит, что распространение света происходит вдоль прямых линий [1]. Другое доказательство дает общеизвестный способ получения изображений светящихся предметов в камере с малым отверстием (камере-обскуре, рис. 2).

Наблюдаются отступления от закона прямолинейного распространения света. Рассмотрим, например, тень от резкого края непрозрачного предмета. Если источник света точечный, то, согласно этому закону, следовало бы ожидать, что на экране получится совершенно резкий переход от света к тени. На самом деле возникает переходная область, в которой освещенность меняется непрерывно и не монотонно: в ней наблюдаются «дифракционные полосы». Аналогичное положение имеет место в камере с малым отверстием. Если отверстие недостаточно мало, то изображение предмета размыто, так как светящаяся точка изображается в виде светлого кружка. Можно было бы ожидать, что при уменьшении отверстия размеры кружка будут уменьшаться, а отчетливость изображения увеличиваться, хотя само изображение и получится менее ярким. На самом деле это оправдывается лишь до известного предела, а при дальнейшем уменьшении отверстия резкость изображения снова начинает ухудшаться. Когда диаметр отверстия $\sim 10^{-3}$ мм, а источник света точечный, получается практически равномерная освещенность экрана.

Приведенные примеры показывают, что свет, подобно звуку, огибает препятствия, стоящие на пути его распространения. Это явление называется *дифракцией*.

Если среда мутная, например туман, то из-за дифракции прямолинейное распространение света сопровождается его *рассеянием в стороны*.

3. Закон независимости световых пучков состоит в том, что *распространение всякого светового пучка в среде совершенно не зависит от того, есть в ней другие пучки света или нет.* Световой пучок, прошедший через какую-либо область пространства, выходит из нее одним и тем же, независимо от того, заполнена она другим светом или не заполнена. Так, изображение на сетчатке глаза не изменится, если свет, образующий это изображение, будет на своем пути проходить через боковые пучки света, не попадающие в глаз. Границы применимости закона независимости световых пучков лучше рассматривать в связи с теорией света (см. § 3, 5 и гл. XI).

---

[1] Протяженный источник ведет себя как совокупность точечных источников, каждому из которых соответствует свое распределение светового поля на экране. В результате наложения таких картин наряду с тенями возникают полутени, т.е. области более или менее плавного перехода от освещенных частей экрана к менее освещенным и совсем темным.

Закон независимости световых пучков необходимо дополнить утверждением, определяющим совместное действие световых пучков при их наложении друг на друга. Оно состоит в том, что *освещенность экрана, создаваемая несколькими световыми пучками, равна сумме освещенностей, создаваемых каждым пучком в отдельности.* Нарушения справедливости этого утверждения имеют место в явлениях *интерференции света.*

4. На основе законов прямолинейного распространения и независимости световых пучков сложилось представление о *световых лучах. В математическом смысле луч есть линия, вдоль которой распространяется свет.* Это — математическая абстракция. *О существовании луча в таком смысле можно говорить лишь постольку, поскольку он входит в состав светового пучка,* содержащего бесконечное множество лучей. Реальное существование имеют не математические лучи и бесконечно тонкие пучки света, а *пучки конечного поперечного сечения,* вырезаемые, например, диафрагмами. Поэтому *под лучом в физическом смысле этого слова мы будем понимать конечный, но достаточно узкий световой пучок, который еще может существовать изолированно от других пучков.* Луч, выделенный какой-либо диафрагмой, не может быть бесконечно длинным, так как из-за дифракции распространение света сопровождается увеличением поперечных размеров луча, т.е. его боковым уширением и размытием. Чем длиннее луч, тем больше это дифракционное уширение. О луче можно говорить только тогда, когда уширение мало по сравнению с поперечными размерами самого луча. Для этого длина луча не должна превышать определенного предела, величина которого тем больше, чем шире луч (см. § 6, формулу (6.17)).

5. Когда луч достигает плоской границы раздела двух прозрачных сред, он частично проходит во вторую среду (преломляется), частично возвращается обратно (отражается). Закон отражения света был известен еще грекам. Он утверждает, что *падающий и отраженный лучи лежат в одной плоскости с нормалью к границе раздела в точке падения* (эта плоскость называется *плоскостью падения*), *причем угол падения φ равен углу отражения φ'* (рис. 3).

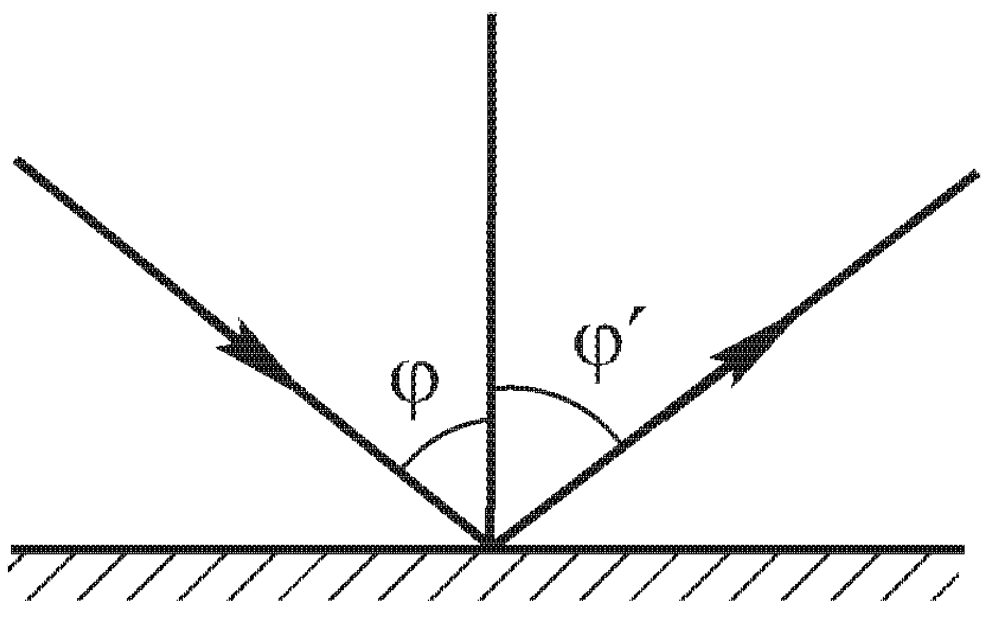


Рис. 3

6. Закон преломления был установлен экспериментально в 1621 г. голландским ученым Снеллиусом (1580–1626) и опубликован только после его смерти. Позднее Декарт (1596–1650) в

1637 г. опубликовал тот же закон, не ссылаясь на Снеллиуса. Знал ли Декарт работы Снеллиуса — этот вопрос остался открытым, хотя он и был предметом многочисленных дискуссий. Декарт получил закон преломления Снеллиуса, пользуясь аналогией между преломлением света и прохождением упругого шара (мяча) через границу раздела воздуха с водой. Его рассуждения были неубедительны и крайне туманны, но окончательный результат, к которому он пришел, оказался верным.

Согласно закону преломления Снеллиуса, *преломленный луч лежит в плоскости падения, причем отношение синуса угла падения* φ (рис. 4) *к синусу угла преломления* ψ *для рассматриваемых сред зависит только от длины световой волны, но не зависит от угла падения, т.е.*

$$\frac{\sin\varphi}{\sin\psi} = n_{21}. \tag{2.1}$$

Постоянная величина n_{21} называется *относительным показателем* или *коэффициентом преломления второй среды относительно первой. Показатель преломления среды относительно вакуума называют абсолютным показателем (коэффициентом) преломления этой среды.* Его будем обозначать через n, снабжая эту букву, если требуется, соответствующими индексами. Например, n_1 — показатель преломления первой, а n_2 — второй сред. Ради краткости величину n обычно называют просто показателем (коэффициентом) преломления среды, т.е. опускают прилагательное «абсолютный».

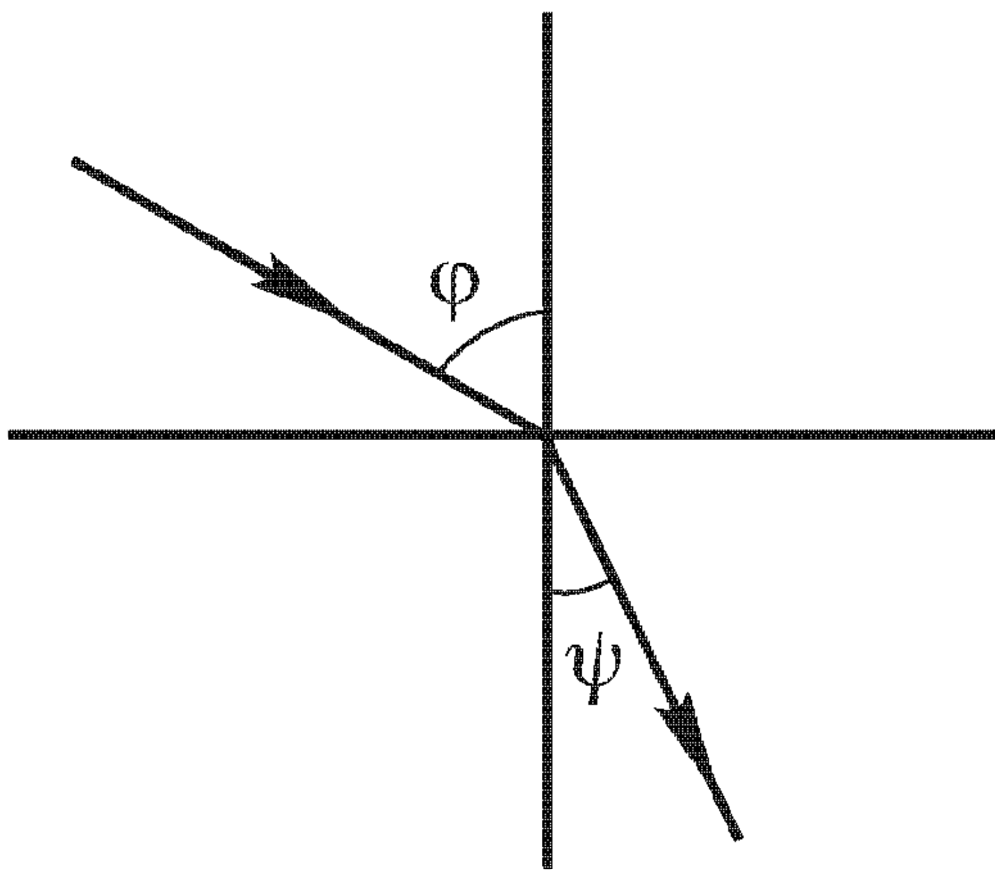


Рис. 4

Относительный показатель преломления n_{21} выражается через абсолютные показатели n_1 и n_2 соотношением

$$n_{21} = n_2/n_1. \tag{2.2}$$

Это соотношение можно получить путем предельного перехода. Пусть световой луч падает из вакуума на плоскопараллельную пластинку с показателем преломления n_1, а затем попадает в среду с показателем преломления n_2 (рис. 5). Для преломления на границах пластинки можно написать

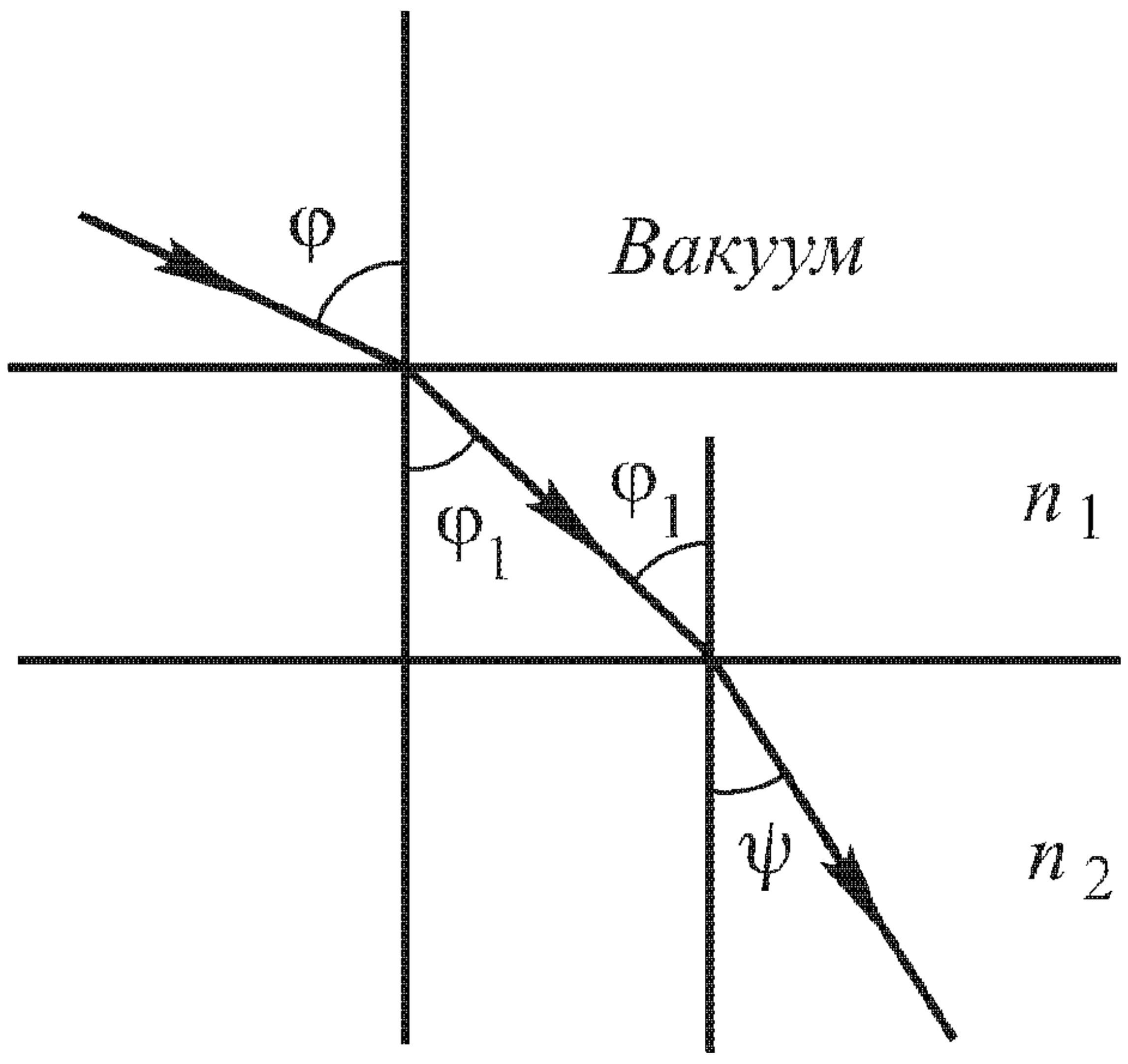


Рис. 5

$$\frac{\sin\varphi}{\sin\varphi_1} = n_1, \qquad \frac{\sin\varphi_1}{\sin\psi} = n_{21}.$$

Перемножая написанные равенства, получим

$$\frac{\sin\varphi}{\sin\psi} = n_{21}n_1. \tag{2.3}$$

Это соотношение справедливо, какова бы ни была толщина пластинки. Оно остается верным и в предельном случае, когда толщина пластинки стремится к нулю. Но тогда свет будет преломляться так, как если бы никакой пластинки вообще не было. Поэтому должно быть $\sin\varphi/\sin\psi = n_2$. Сравнение этого результата с предыдущим и приводит к соотношению (2.2).

Слабая сторона приведенного рассуждения состоит в следующем. Показатель преломления есть *макроскопическая характеристика среды.* Когда толщина пластинки, разделяющей среды 1 и 2, становится порядка атомных размеров, ее уже нельзя рассматривать как непрерывную среду, так что понятие показателя преломления теряет смысл. Однако окончательный результат (2.2) остается верным. Он подтверждается опытом и в дальнейшем при рассмотрении теории отражения и преломления света будет выведен с различных точек зрения (см. § 3, 64).

С учетом соотношения (2.2) закон преломления можно записать в симметричной форме:

$$n_1\sin\varphi = n_2\sin\psi. \tag{2.4}$$

Из формулы (2.2) следует также:

$$n_{21} = 1/n_{12}. \tag{2.5}$$

7. Если $n_{21} < 1$, то может оказаться, что величина $\sin\psi$, формально вычисленная по формуле (2.1), начнет превосходить единицу, т.е. $\varphi/n_{21} > 1$. Соответствующего угла преломления не существует. Поэтому преломленный луч не возникает, а свет отражается полностью. Это явление называется *полным отражением.* Угол падения, при котором оно возникает, определяется условием $\varphi \geqslant \varphi_0$, причем

$$\sin\varphi_0 = n_{21}. \tag{2.6}$$

Величина φ_0 называется *предельным углом полного отражения.* Полное отражение будет исследовано в § 66.

8. Если поверхность тела, на которую падают световые лучи, не плоская, а кривая, то ее можно мысленно разбить на малые площадки, считая каждую из них плоской. Тогда ход лучей можно найти по законам отражения и преломления, изложенным выше, а затем выполнить предельный переход к гладкой поверхности, устремляя к нулю размеры каждой площадки. Однако этот прием применим только тогда, когда кривизна поверхности не превышает некоторого предела, так как в противном случае начинают проявляться отступления от законов правильного отражения и преломления и наступает дифракция.

Шероховатые поверхности дают не правильное, а *рассеянное*, или *диффузное*, отражение и преломление света. Только благодаря этому поверхность тела становится видимой. Абсолютно зеркальная поверхность невидима, видны только отраженные от нее лучи, попадающие от источников света, расположенных вне зеркала, т.е. видны только сами эти источники света.

ЗАДАЧИ

1. Два плоских зеркала *1* и *2* наклонены друг к другу, образуя двугранный угол α (рис. 6). Падающий луч, лежащий в плоскости, перпендикулярной к ребру двугранного угла, отражается сначала от одного, а затем от другого зеркала. Показать, что в результате этих двух отражений луч отклоняется на угол δ, величина которого не зависит от направления падающего света. Вычислить угол δ.

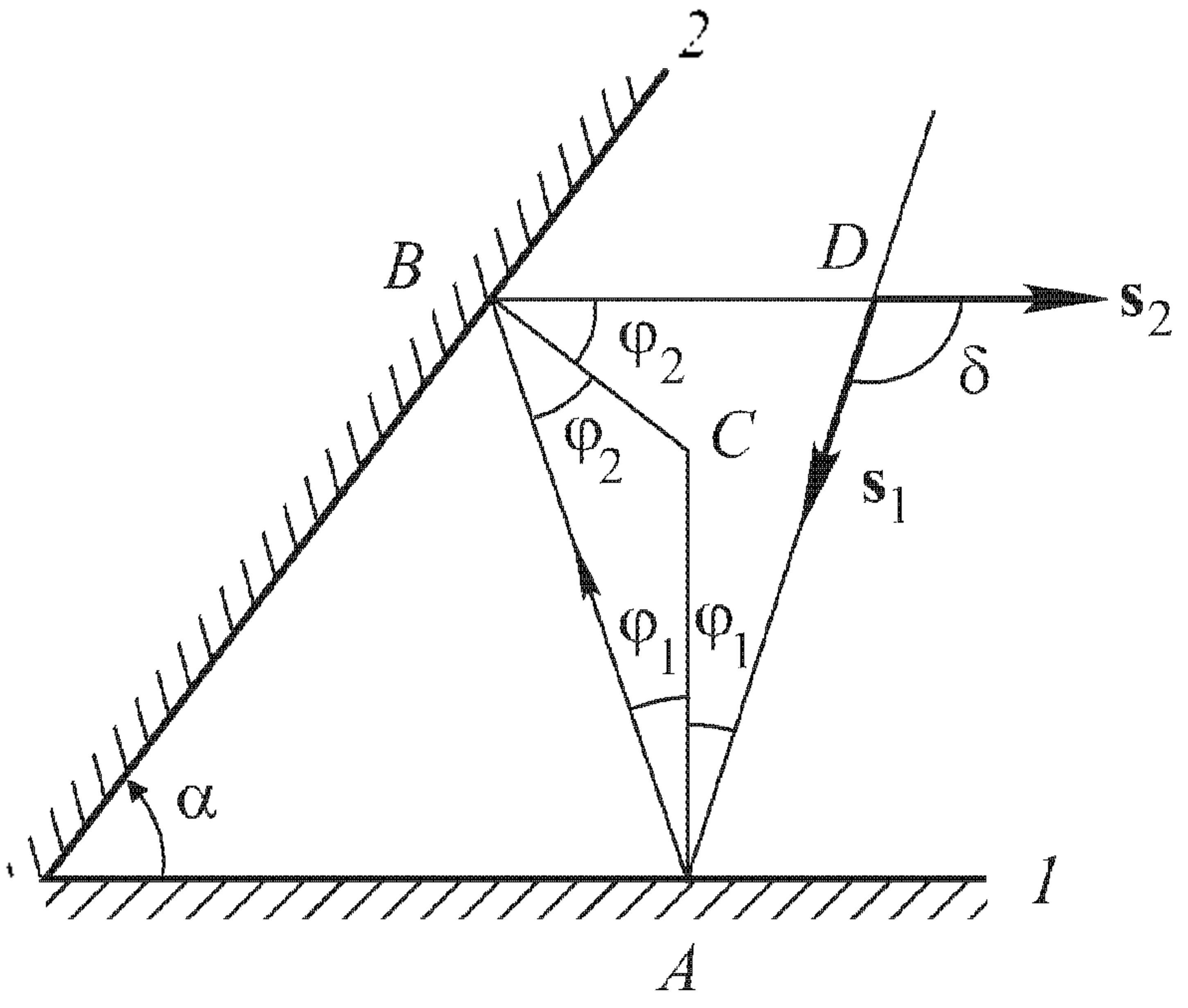


Рис. 6

Решение. Как видно из треугольника ABD, $\delta = 2(\varphi_1 + \varphi_2)$, а из треугольника ABC $\varphi_1 + \varphi_2 = \alpha$. Поэтому $\delta = 2\alpha$. Результат справедлив при любом α и любом направлении падающего света, если только угол α отсчитывать от зеркала *1*, вращая его по кратчайшему пути к зеркалу *2*, а угол δ — от направления падающего луча s_1 к направлению выходящего луча s_2, производя вращение в том же направлении.

2. Показать, что луч света, последовательно отражающийся от трех взаимно перпендикулярных зеркал, меняет свое направление на противоположное.

Решение. Пусть $\mathbf{s}_0$, $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$ — единичные векторы падающего и отраженных лучей, а $\mathbf{N}_1$, $\mathbf{N}_2$, $\mathbf{N}_3$ — единичные нормали к отражающим плоскостям зеркал. Тогда

$$\mathbf{s}_1 - \mathbf{s}_0 = -2(\mathbf{N}_1\mathbf{s}_0)\mathbf{N}_1,$$
$$\mathbf{s}_2 - \mathbf{s}_1 = -2(\mathbf{N}_2\mathbf{s}_1)\mathbf{N}_2,$$
$$\mathbf{s}_3 - \mathbf{s}_2 = -2(\mathbf{N}_3\mathbf{s}_2)\mathbf{N}_3.$$

Из первого уравнения скалярным умножением на $\mathbf{N}_2$ и $\mathbf{N}_3$ получаем

$$(\mathbf{N}_2\mathbf{s}_1) = (\mathbf{N}_2\mathbf{s}_0), \quad (\mathbf{N}_3\mathbf{s}_1) = (\mathbf{N}_3\mathbf{s}_0).$$

Аналогично, скалярное умножение второго уравнения на $\mathbf{N}_3$ дает

$$(\mathbf{N}_3\mathbf{s}_2) = (\mathbf{N}_3\mathbf{s}_1) = (\mathbf{N}_3\mathbf{s}_0).$$

Почленно складывая все три уравнения и учитывая найденные соотношения, получим

$$\mathbf{s}_3 - \mathbf{s}_0 = -2(\mathbf{N}_1\mathbf{s}_0)\mathbf{N}_1 - 2(\mathbf{N}_2\mathbf{s}_0)\mathbf{N}_2 - 2(\mathbf{N}_3\mathbf{s}_0)\mathbf{N}_3 = -2\mathbf{s}_0,$$

откуда $\mathbf{s}_3 = -\mathbf{s}_0$.

Полученный результат лежит в основе устройства уголкового отражателя, применяющегося для изменения направления распространения света на противоположное. Уголковый отражатель можно получить, отсекая от стеклянного куба с посеребренными гранями трехгранный угол плоскостью, перпендикулярной к пространственной диагонали куба. Всякий луч, вступивший внутрь так полученной пирамиды через ее основание, испытав отражения от трех ее боковых граней, выйдет через то же основание, изменив свое направление на противоположное. Действительно, преломление на основании пирамиды, испытываемое падающим лучом, не играет роли, поскольку оно полностью компенсируется преломлением на том же основании при выходе луча из пирамиды.

3. Луч света проходит через ряд однородных сред, разграниченных плоскостями, параллельными между собой. Показать, что направление луча в последней среде (если луч проникает в нее) зависит только от угла падения и от показателей преломления первой и последней сред. В частности, если показатели преломления первой и последней сред одинаковы, то луч войдет в последнюю среду параллельно тому направлению, которое он имел в первой среде.

4. В преломляющей призме (рис. 7) $\angle C = \angle D$, $\angle A = \alpha$, $\angle B = 2\alpha$. Световой луч вступает в призму через грань BD, находясь в ее главном сечении, т.е. в плоскости, перпендикулярной к преломляющим ребрам призмы, а затем последовательно отражается от граней AC и AD, выходя наружу через грань BC. Показать, что угол δ отклонения вышедшего луча от исходного направления не зависит от угла падения. Вычислить угол δ.

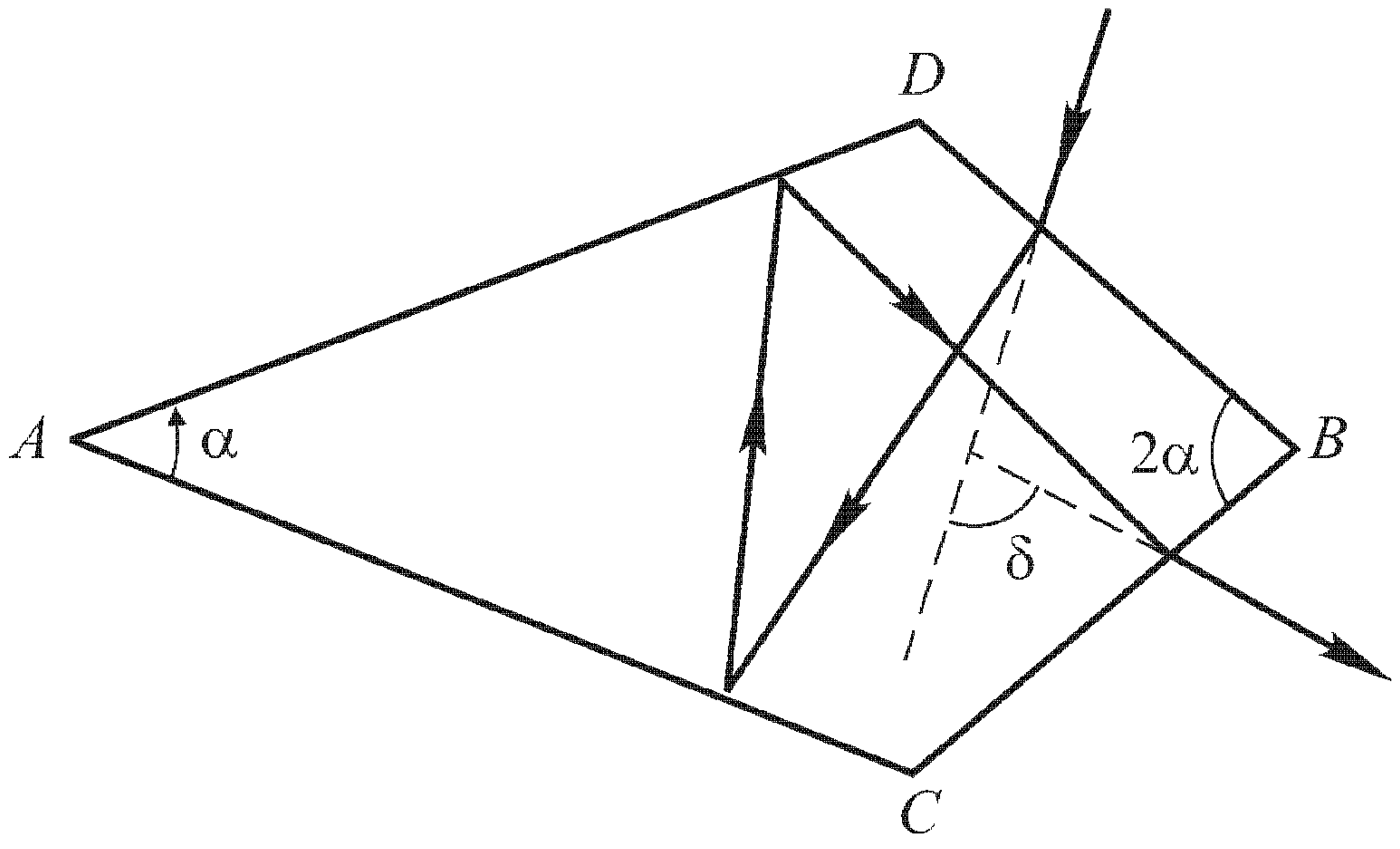


Рис. 7

Ответ. $\delta = 2\alpha$.

5. Исследовать преломление светового луча в главном сечении трехгранной призмы. Определить угол отклонения светового луча δ от исходного направления и его наименьшее значение $\delta_{\text{мин}}$.

Решение. Как видно из треугольника CDE (рис. 8),

$$\delta = (\varphi_1 - \psi_1) + (\varphi_2 - \psi_2).$$

Далее, из четырехугольника $ACFD$, у которого углы C и D — прямые, имеем: $A + F = \pi$, а из треугольника CDF $(\psi_1 + \psi_2) + F = \pi$, так что $\psi_1 + \psi_2 = A = \text{const}$, где A — преломляющий угол призмы. Следовательно,

$$\delta = \varphi_1 + \varphi_2 - A. \tag{2.7}$$

Для нахождения минимума угла отклонения δ удобно за независимую переменную принять угол преломления ψ_1, так как при этом будет достигнута симметрия и упрощение выкладок. Для первой производной получаем

$$\frac{d\delta}{d\psi_1}=\frac{d\varphi_1}{d\psi_1}+\frac{d\varphi_2}{d\psi_1}=\frac{d\varphi_1}{d\psi_1}-\frac{d\varphi_2}{d\psi_2},$$

или, на основании закона преломления,

$$\frac{d\delta}{d\psi_1}=n\left(\frac{\cos\psi_1}{\cos\varphi_1}-\frac{\cos\psi_2}{\cos\varphi_2}\right). \tag{2.8}$$

Аналогично, для второй производной

$$\frac{d^2\delta}{d\psi_1^2}=n\left(\frac{d}{d\psi_1}\frac{\cos\psi_1}{\cos\varphi_1}-\frac{d}{d\psi_1}\frac{\cos\psi_2}{\cos\varphi_2}\right)=n\left(\frac{d}{d\psi_1}\frac{\cos\psi_1}{\cos\varphi_1}+\frac{d}{d\psi_2}\frac{\cos\psi_2}{\cos\varphi_2}\right)=$$

$$=n\sum_{i=1}^{2}\frac{\sin\psi_i}{\cos^3\psi_i}(n^2\cos^2\psi_i-\cos^2\varphi_i).$$

Если $n>1$, то $d^2\delta/d\psi_1^2>0$, а потому кривая $\delta=\delta(\psi_1)$ во всех точках обращена выпуклостью вниз. Отсюда следует, что угол δ достигает минимума при $\varphi_1=\varphi_2$, т.е. при симметричном ходе луча через призму. Других минимумов (и вообще экстремумов) быть не может.

При симметричном ходе луча $\varphi={}^1\!/_2(A+\delta_{\text{мин}})$, $\psi={}^1\!/_2A$, и следовательно,

$$n=\frac{\sin[(A+\delta_{\text{мин}})/2]}{\sin(A/2)}. \tag{2.9}$$

На этой формуле основан удобный метод измерения показателя преломления.

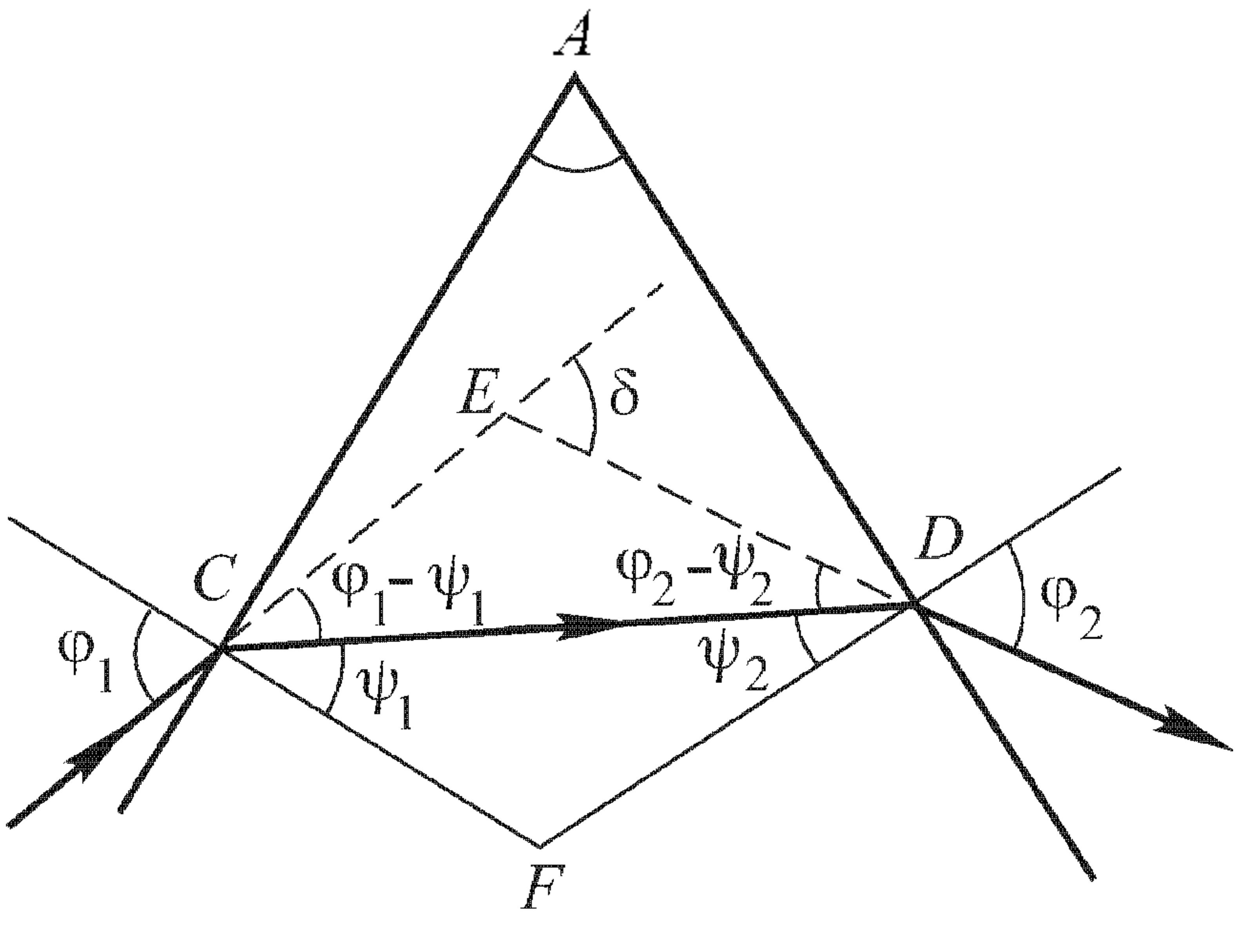


Рис. 8

6. Световой луч падает на боковую грань призмы под малым углом. Преломляющий угол призмы A мал (рис. 8). Вычислить угол отклонения луча δ, ограничиваясь членами первой степени по A. Вычислить также угол наименьшего отклонения $\delta_{\text{мин}}$ с точностью до членов порядка A^3 включительно.

Решение. $\delta=(\varphi_1-\psi_1)+(\varphi_2-\psi_2)\approx(n\psi_1-\psi_1)+(n\psi_2-\psi_2)=$ $=(n-1)(\psi_1+\psi_2)$, или, на основании соотношения $\psi_1+\psi_2=A$,

$$\delta=(n-1)A. \tag{2.10}$$

В рассматриваемом приближении угол δ не зависит от угла падения. В высших приближениях это уже не так. В третьем приближении из формулы (2.9), разлагая $\delta_{\text{мин}}$ в ряд, находим

$$\delta_{\text{мин}}=(n-1)A\left[1+\frac{n(n+1)}{24}A^2\right]. \tag{2.11}$$

7. Удобный метод измерения показателя преломления твердых тел состоит в следующем. Плоскопараллельная пластинка из исследуемого вещества рассматривается в микроскоп. Сначала микроскоп устанавливается для наблюдения верхней поверхности пластинки. Затем смещают тубус микроскопа вниз, пока не будет отчетливо видна нижняя поверхность пластинки. (Для удобства наблюдения на поверхностях пластинки можно сделать метки.) Показать, что если смещение тубуса равно h, а толщина пластинки l, то показатель преломления пластинки можно найти по формуле

$$n = l/h.$$

Метод годится и для измерения показателей преломления жидкостей.

§ 3. Эволюция представлений о природе света

1. Пифагор (около 580–500 до н.э.) считал, что предметы становятся видимыми благодаря мельчайшим частицам, испускаемым ими и попадающим в глаз наблюдателя. Декарт полагал, что свет — это сжатие, распространяющееся в идеальной упругой среде (эфире), заполняющей мировое пространство и промежутки между частицами тел. Однако последовательно провести эту точку зрения Декарт не смог, при выводе законов отражения и преломления он пользовался представлением о свете как о потоке частиц. Гук (1635–1703) также считал, что свет представляет собой импульсы сжатия, распространяющиеся мгновенно или с очень большими скоростями. (Скорость света была определена только в 1676 г. Олафом Рёмером (1644–1710) из наблюдений затмений спутников Юпитера.) Несколькими годами ранее Гука чешский монах Марци (1595–1667) и итальянский монах Гримальди (1618–1663) также пришли к мысли, что свет представляет собою быстро распространяющиеся волны. Подобные отрывочные высказывания о природе света были усовершенствованы и развиты в более систематические теории Исааком Ньютоном (1643–1727), с одной стороны, и Христианом Гюйгенсом (1629–1695), с другой.

Основные оптические работы Ньютона докладывались в Лондонском Королевском обществе (1671–1675 гг.) и были напечатаны в трудах этого общества. Содержание этих сообщений, наряду с другими оптическими исследованиями Ньютона, было заимствовано из его «Лекций по оптике» [1]), которые он читал в Кембридже в 1669–1671 гг. Спустя много лет (в 1704 г.) вышла знаменитая книга Ньютона «Оптика» [2]). В оптике Ньютон сосредоточил основное внимание на экспериментальных исследованиях и стремился при изложении этого, как и других разделов физики, отделить достоверные факты от сомнительных гипотез. Экспериментальные исследования Ньютона по оптике вплоть до XIX века не имели себе равных. Его «Оптика» служила основным источником, из которого черпали материал все учебники того времени («Лекции по оптике» были опубликованы после смерти Ньютона и по ряду причин остались почти неизвестными). Что касается природы света, то к этому вопросу Ньютон подходил значительно более осторожно, чем все остальные ученые, включая Гюйгенса. Он предпочитал с самого начала строить чисто феноменологическую теорию, освобожденную от произвольных гипотез. Все, что относится к физической природе света, он выделил в особый раздел, содержащий 31 «вопрос», которым заканчивается его «Оптика». Здесь после тщательного рассмотрения различных исключающих друг друга возможностей Ньютон пришел к заключению, что факты говорят больше в пользу корпускулярной теории, чем волновой. Поэтому он, правда не без колебаний и сомнений, отдал предпочтение корпускулярной теории.

---

[1]) *Исаак Ньютон.* Лекции по оптике. Перевод, комментарии и редакция академика С.И. Вавилова. — М.: Издательство АН СССР, 1946.

[2]) *И. Ньютон.* Оптика или трактат об отражениях, преломлениях, изгибаниях и цветах света. Перевод с примечаниями С.И. Вавилова. — М.: Гостехиздат, 1954.

Гюйгенс в своих высказываниях о природе света не был столь сдержанным и осторожным, как Ньютон. Он не сомневался в волновой природе света. Свои воззрения он изложил на заседании Парижской Академии наук (1678 г.), а затем более полно в сочинении «Трактат о свете» [1]), опубликованном в 1690 г. В этом сочинении впервые в истории оптики была сделана попытка изложить с единой — волновой — точки зрения распространение, отражение, преломление и двойное лучепреломление света.

2. Перейдем к более подробной характеристике корпускулярной и волновой теорий света.

Согласно корпускулярной теории, свет состоит из мельчайших частиц, или корпускул, испускаемых светящимися телами. С этой точки зрения прямолинейное распространение света сводится к закону инерции. Для истолкования закона независимости световых пучков надо было ввести предположение, что средние расстояния между корпускулами в световых пучках настолько велики, что корпускулы практически не взаимодействуют между собой; случаи сближения, в которых проявляется такое взаимодействие, крайне редки и при существующей точности эксперимента ускользают от наблюдения.

Отражение и преломление света корпускулярная теория объясняла силами притяжения и отталкивания, действующими на световые корпускулы в очень тонком приграничном слое вблизи границы раздела сред, на которую падает свет. Внутри приграничного слоя путь световой корпускулы искривляется, по выходе из него корпускула движется снова прямолинейно и равномерно, но уже в другом направлении. Если корпускула отражается, то она возвращается в первую среду с прежним значением скорости. Если же корпускула проходит во вторую среду, то величина ее скорости изменяется (рисунки 9 и 10). Ввиду тонкости пограничного слоя явление воспринимается так, как если бы на границе сред происходил резкий излом траектории корпускулы (как показано на рисунках 9 и 10 штриховыми линиями).

---

[1]) *Х. Гюйгенс.* Трактат о свете. — М.; Л.: ОНТИ, 1935.

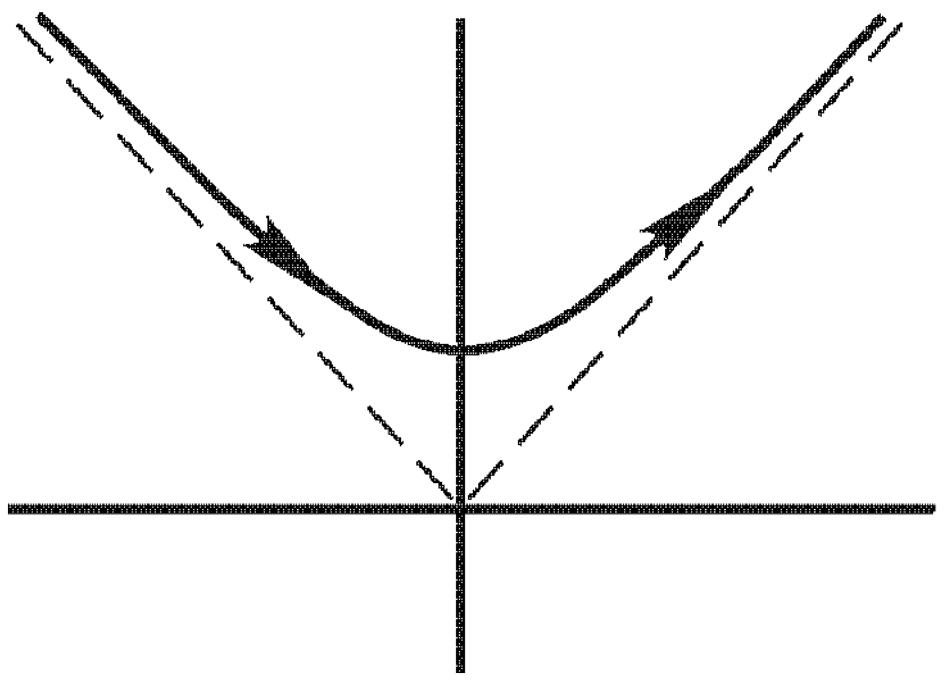


Рис. 9

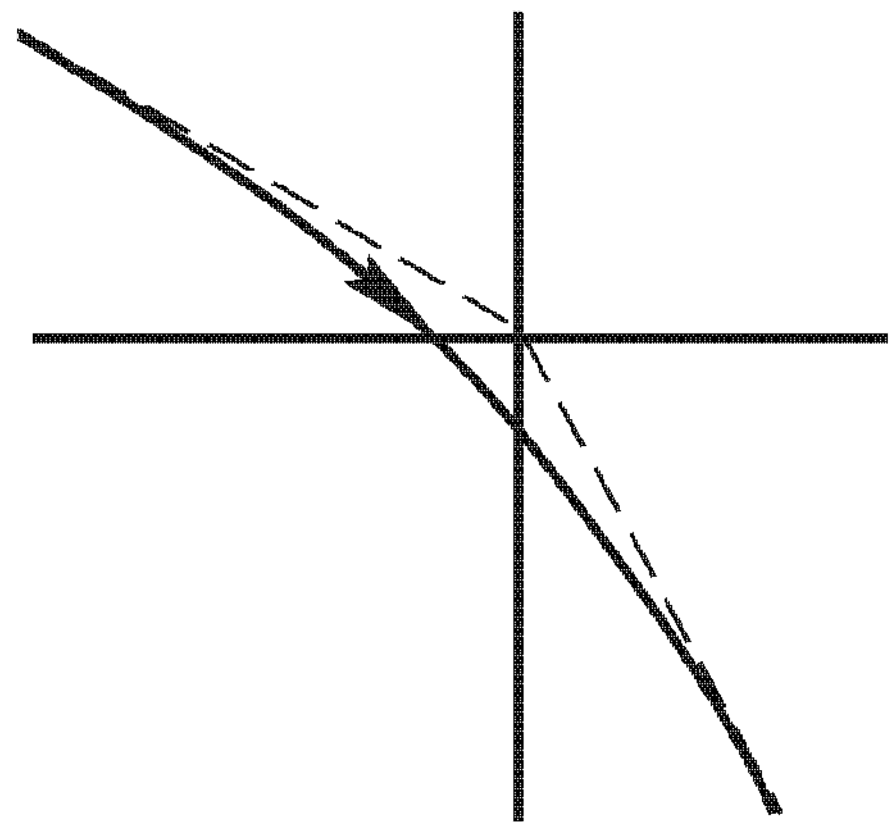


Рис. 10

Для вывода количественных законов отражения и преломления света надо было ввести предположение, что силы, действующие на световую корпускулу в пограничном слое, *нормальны к границе раздела сред.* Такие силы меняют только нормальные скорости световой корпускулы, оставляя касательные без изменения. Рассмотрим, например, преломление света. Обозначим через v_1 скорость световой корпускулы в первой среде, а через v_2 — во второй. В силу равенства касательных составляющих этих скоростей $v_1 \sin\varphi = v_2 \sin\psi$ (рис. 11), или

$$\frac{\sin\varphi}{\sin\psi} = \frac{v_2}{v_1}. \tag{3.1}$$

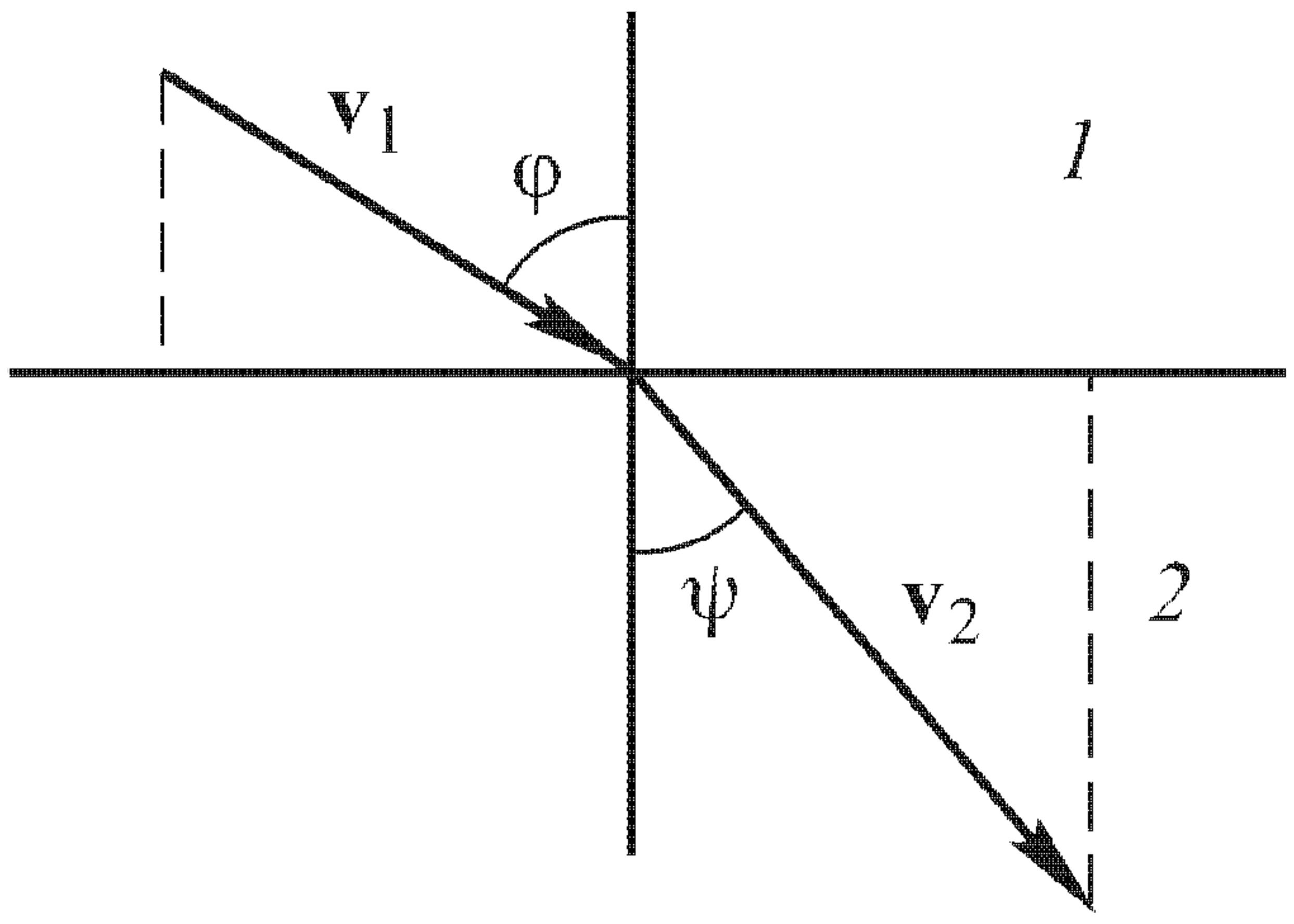


Рис. 11

Эта формула была получена Ньютоном. *Из нее следует, что в сильнее преломляющих средах скорость света должна быть больше, чем в менее преломляющих.* Однако чтобы из формулы (3.1) вывести закон преломления Снеллиуса, необходимо добавочное предположение, что отношение скоростей света v_2 и v_1 для световых корпускул одного и того же типа постоянно, т.е. определяется только свойствами сред *1* и *2*, в которых корпускулы движутся, но не зависит от того, каким путем они туда попали. Если использовать экспериментальный факт, что в вакууме скорость света одна и та же, то для выполнения этого условия необходимо потребовать, чтобы все силы, действующие на световые корпускулы в приграничных слоях, были *потенциальны*. (Заметим, что цвета света тогда можно объяснить, например, различными размерами корпускул.) Однако то обстоятельство, что скорость световой корпускулы зависит только от среды, в которую она испускается, но не зависит от способа испускания, представляется малопонятным с точки зрения классической механики. В этом отношении корпускулярная теория уступает волновой теории, поскольку в последней скорость света, естественно, есть характеристика *только самой среды*, в которой он распространяется.

При справедливости введенных предположений показатель преломления среды *2* относительно среды *1* определяется выражением

$$n_{21} = v_2/v_1. \tag{3.2}$$

В частности, для абсолютного показателя преломления получаем

$$n = v/c, \tag{3.3}$$

где c — скорость света в вакууме. Следовательно, $n_1 = v_1/c$, $n_2 = v_2/c$, а потому

$$n_{21} = n_2/n_1, \tag{3.4}$$

что совпадает с формулой (2.2). Однако это обстоятельство нельзя рассматривать как аргумент в пользу корпускулярной теории, так как такое же соотношение следует и из волновой теории.

3. Согласно *волновой теории*, свет представляет собой *волны*, распространяющиеся в гипотетической всепроникающей среде, — *мировом* или *световом эфире* — заполняющей все мировое пространство и промежутки между мельчайшими частицами тел. Если колебания частиц эфира малы, то уравнения, описывающие распространение волн, будут *линейны и однородны*. В этом случае справедлив *принцип суперпозиции волн*, являющийся в волновой теории математическим выражением закона независимости световых пучков.

Значительно сложнее обстояло дело с объяснением прямолинейного распространения света. Гюйгенс применил к этой проблеме принцип, названный позднее его именем. Если камень бросить в воду, то от места падения, как из центра, побегут по поверхности воды круговые волны, или возмущения. Этот процесс продолжается и после того, как камень упадет на дно, т.е. исчезнет причина, породившая первоначальное возмущение. Отсюда следует, что непосредственной причиной распространяющегося волнового процесса является не камень, а то первоначальное возмущение движения воды, которое он вызвал. Подобные рассуждения навели Гюйгенса на мысль, что *каждая точка волнового* *возмущения является источником сферических волн*, распространяющихся от нее во все стороны. Они получили название *вторичных* или *элементарных волн Гюйгенса*. Результирующее волновое возмущение можно рассматривать как наложение вторичных волн. В этом и состоит знаменитый *принцип Гюйгенса*. Он справедлив не только для световых, но и для *любых* волновых процессов и в приведенной формулировке полностью сохранил свое значение до настоящего времени. Его значение далеко выходит за рамки сравнительно узкой задачи о прямолинейном распространении света. Действительно, принцип Гюйгенса указывает на принципиальную возможность определения волнового возмущения во всем пространстве, если известны все вторичные волны, порожденные в более ранний момент времени.

Но Гюйгенс не рассматривал задачу в столь общей постановке. Он дополнил формулировку своего принципа, приведенную выше, замечанием, что отдельные вторичные волны слишком слабы и что заметное световое действие они производят только на их *огибающей*. После этого дополнения принцип Гюйгенса потерял свою общность и превратился лишь в приближенный *геометрический рецепт* для построения волновых фронтов, т.е. поверхностей, до которых дошло световое возмущение. Он стал непригоден для количественного расчета светового поля, в частности распределения интенсивности последнего в пространстве.

4. Обратимся теперь к объяснению прямолинейного распространения света, данному Гюйгенсом. Пусть точечный источник света S (рис. 12), расположенный перед непрозрачным экраном с отверстием AB, излучает сферическую волну. Если в момент времени t волна дойдет до краев отверстия AB, то оно вырежет из волнового фронта сферический участок ACB. Построим из каждой точки этого участка, как из центра, вторичные волны радиусом $c\,dt$. Огибающая всех этих вторичных волн оборвется на краях сферического участка A_1B_1, за пределы которого проникнут только отдельные вторичные волны. Действие таких волн, согласно предположению Гюйгенса, пренебрежимо мало, а потому волновой фронт возмущения в момент $t + dt$ ограничится только сферическим участком A_1B_1. Взяв этот фронт за исходный, можно таким же построением найти волновой фронт A_2B_2 в более поздний момент времени, и т.д. Построение показывает, что волновое возмущение будет резко обрываться на поверхности телесного угла с вершиной S, вырезаемого

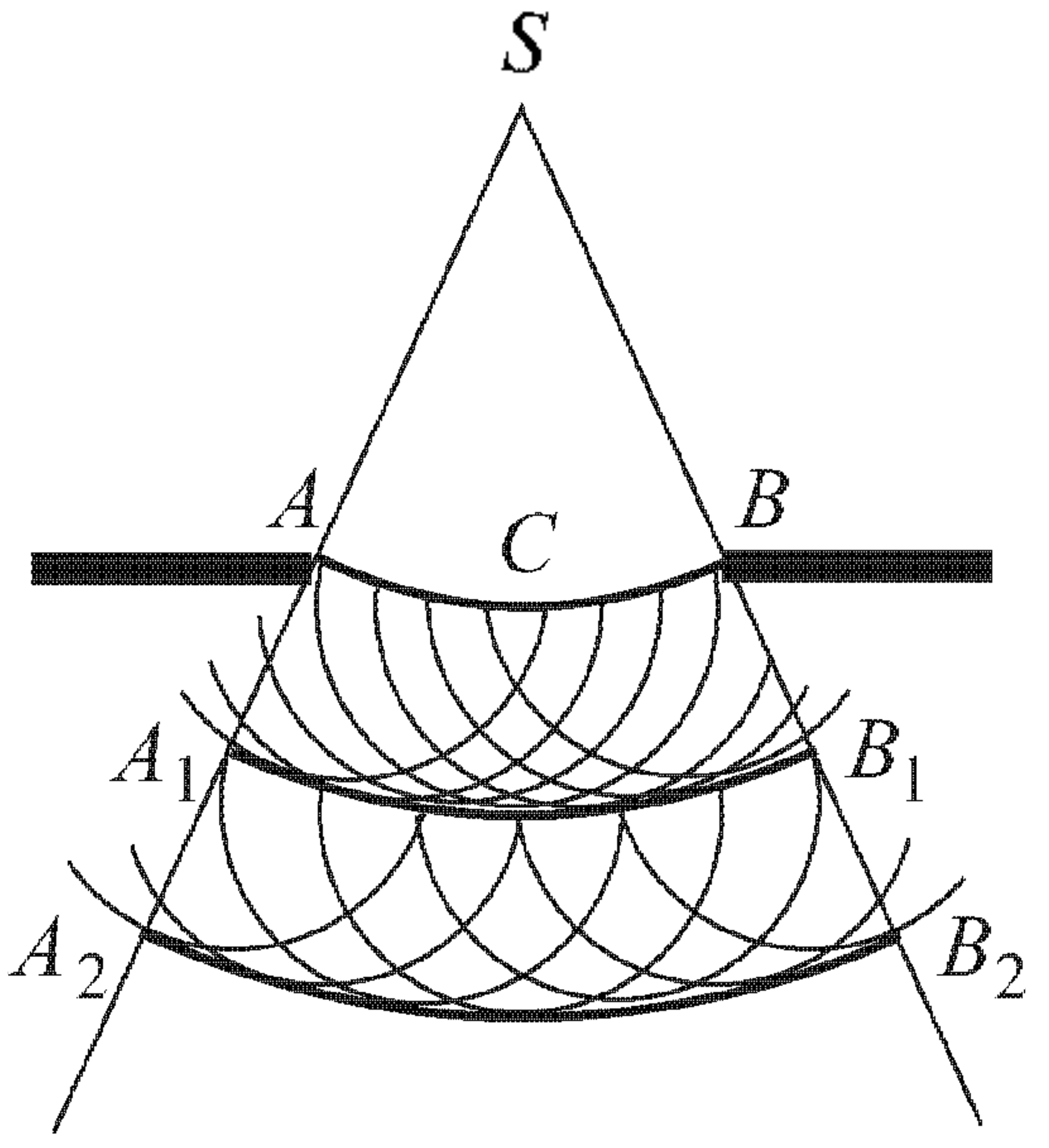


Рис. 12

диафрагмой AB. А это и означает прямолинейное распространение света.

Рассуждение Гюйгенса лишено доказательной силы. В нем доказывается то, что по существу содержится в исходных предположениях. А само построение Гюйгенса есть только *один из возможных рецептов построения волнового фронта*, согласующийся с представлением о распространении света вдоль лучей. Непонятно, почему объяснение Гюйгенса применимо к световым, но не применимо к звуковым волнам, которые, как известно, огибают препятствия, стоящие на их пути.

Естественно, что такое «объяснение» не могло удовлетворить Ньютона. Кроме того, Ньютон, как и Гюйгенс, не видел, как в рамках волновой теории можно объяснить поляризацию света, открытую Гюйгенсом в двойном лучепреломлении. Это действительно невозможно для продольных волн, какими, согласно Гюйгенсу, является свет. Только такие волны и были известны в физике того времени. Представления о поперечных волнах еще не существовало. Эти трудности казались Ньютону настолько существенными, что его симпатии оказались на стороне корпускулярной теории.

Проблема прямолинейного распространения света есть *частный случай проблемы дифракции* и может быть решена до конца только в рамках последней. Дифракция света была открыта Гримальди и независимо от него несколько позднее Гуком. Ньютон много занимался экспериментальными исследованиями дифракции света. Но Гюйгенс в «Трактате о свете» почему-то полностью обошел молчанием это явление. Кроме того, ему осталась неизвестной периодичность световых процессов (в отличие от Ньютона, который первый подметил ее). Гюйгенс писал, что свет, подобно звуку, распространяется *сферическими поверхностями*, и именно такие поверхности называл *волнами*. Он специально подчеркивал, что удары, возбуждающие световые возмущения в центрах волн, совершаются совершенно беспорядочно, а потому не следует думать, что сами волны следуют друг за другом на равных расстояниях. В этом отношении высказывания Гюйгенса примыкают к более ранним представлениям Декарта и Гука. Понятие длины волны нигде не встречается в теории Гюйгенса, а без этого невозможно установить, при каких условиях (приближенно) справедлив закон прямолинейного распространения света.

5. Перейдем к вопросу об отражении и преломлении света в волновой теории Гюйгенса. Пусть плоская волна падает на плоскую границу раздела двух сред. В некоторый момент в точке A волновой фронт AB достигнет границы раздела сред (рис. 13). В этот момент из A начнут распространяться вторичные волны Гюйгенса: одна в первую, другая во вторую среду. В точках E

и D аналогичные волны возникнут несколько позднее. Согласно принципу Гюйгенса, от наложения таких вторичных волн в первой среде образуется *отраженная*, а во второй — *преломленная* волны. Огибающая вторичных волн во второй среде FGD есть плоскость, определяющая волновой фронт преломленной волны. Аналогично строится волновой фронт и отраженной волны (на рис. 13 он не указан). Таково объяснение отражения и преломления света в волновой теории Гюйгенса. Оно не раскрывает детальный механизм возникновения отраженной и преломленной волн. Для этого надо было бы явно использовать физическую природу световых волн, о которой в эфирной теории Гюйгенса, в сущности, ничего не говорится.

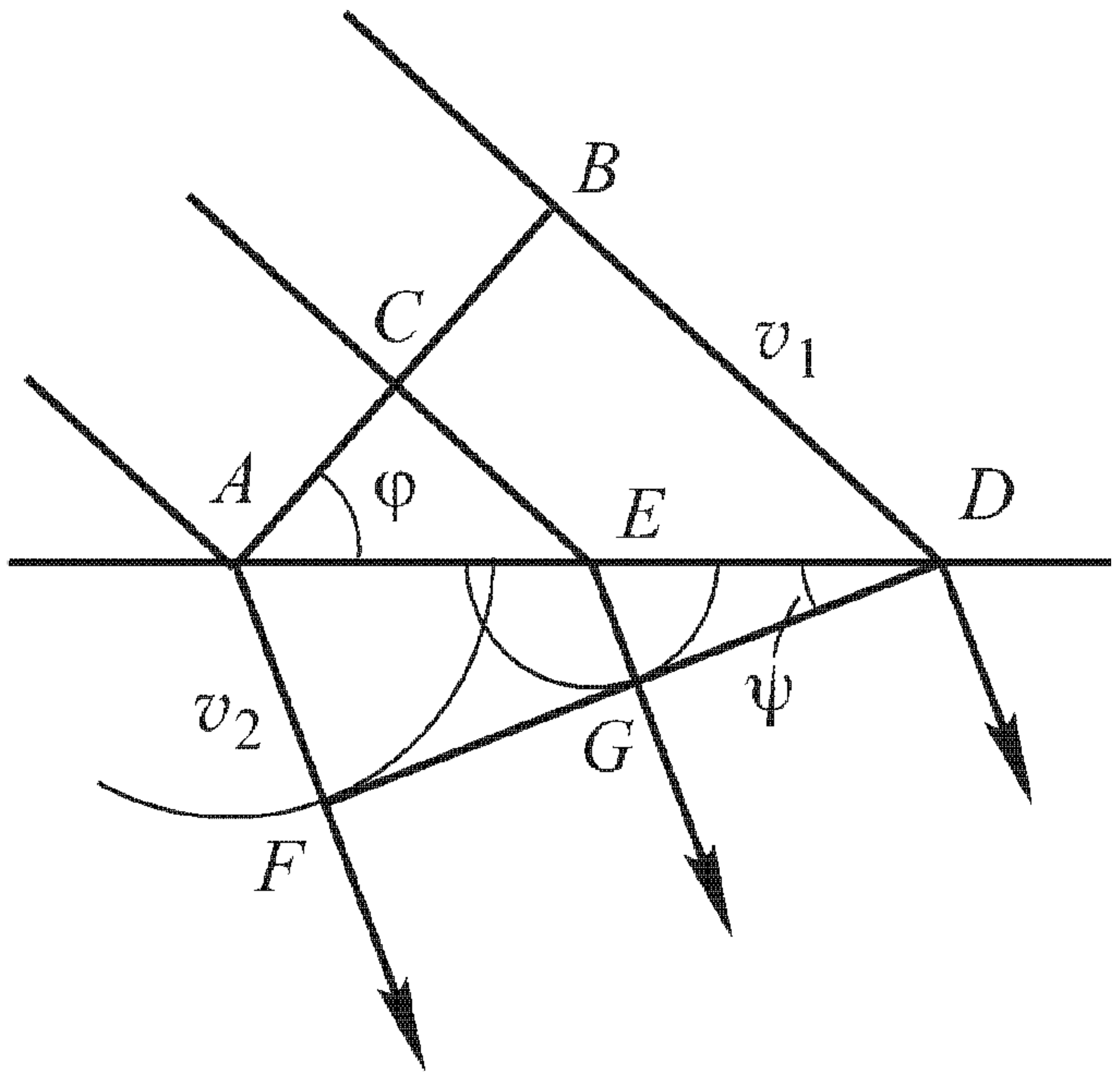


Рис. 13

Геометрические законы отражения и преломления, однако, *совершенно не зависят* от физической природы волн и от конкретного механизма отражения и преломления. Они одинаковы в любой волновой теории. Действительно, падающая волна возбуждает возмущение, бегущее вдоль границы раздела со скоростью $AD = BD/\sin\varphi = v_1/\sin\varphi$ (если воспользоваться надлежащими единицами), где v_1 — скорость света в первой среде. Но отраженная и преломленная волны порождаются падающей волной и поэтому бегут вместе с ней вдоль границы раздела с той же скоростью. Следовательно, можно написать также, что $AD = v_1/\sin\varphi' = v_2/\sin\psi$, где v_2 — скорость света во второй среде, φ' — угол наклона фронта отраженной волны к границе раздела сред (не показанный на рис. 13). В результате получается

$$\frac{\sin\varphi}{v_1} = \frac{\sin\varphi'}{v_1} = \frac{\sin\psi}{v_2}. \tag{3.5}$$

Этими соотношениями определяются *направления фронтов* отраженной и преломленной волн. А так как в плоской волне световые лучи перпендикулярны к волновым фронтам, то те же соотношения определяют также *направления отраженных и преломленных лучей*. Легко видеть, что φ есть угол падения, φ' — угол отражения, ψ — угол преломления. Из (3.5) следует, что $\varphi = \varphi'$ (закон отражения) и

$$\frac{\sin\varphi}{\sin\psi} = \frac{v_1}{v_2} \qquad \text{(закон преломления).} \tag{3.6}$$

В противоположность корпускулярной теории (см. (3.1)), *волновая теория приводит к заключению, что скорость света в более преломляющей среде меньше, чем в менее преломляющей.* И это заключение справедливо независимо от того, какова физическая природа световых волн.

Для относительного показателя преломления волновая теория дает

$$n_{21} = v_1/v_2, \tag{3.7}$$

а для абсолютного

$$n = c/v, \tag{3.8}$$

т.е. выражения, обратные соответствующим выражениям в корпускулярной теории (3.2) и (3.3). Из (3.7) и (3.8) снова получается формула (3.4), которая, следовательно, справедлива как в корпускулярной, так и в волновой теориях.

6. В вопросе об отражении и преломлении света между корпускулярной и волновой теориями имеется и другое различие. Опыт показывает, что световой пучок, падающий на границу раздела сред, разделяется на два пучка: отраженный и преломленный (исключением является случай полного отражения). В волновой теории объяснение этого факта не встречает затруднений. Но как истолковать его с точки зрения корпускулярной? Корпускула всегда ведет себя как целое. Она не может разделиться на две части, из которых одна отражается, а другая проходит во вторую среду. В противном случае падающий, отраженный и прошедший свет был бы окрашен различно. Для преодоления этой трудности Ньютон ввел идею о *приступах легкого отражения и легкого прохождения*, в которых периодически может находиться световая корпускула. Если корпускула подходит к границе сред в приступе легкого отражения, то она отталкивается и отражается. Если же это случится в приступе легкого прохождения, то произойдет притяжение и прохождение корпускулы во вторую среду.

Идея «приступов», как заметил Я.И. Френкель (1894–1952), напоминает современные представления о *световых квантах — фотонах.* Согласно этим представлениям, отражение и прохождение фотонов через границу раздела сред управляется *статистическими законами*: существует определенная вероятность, что фотон отразится, и определенная вероятность, что он пройдет во вторую среду. Сам Ньютон, конечно, не раскрыл физический механизм приступов легкого отражения и легкого прохождения. Однако он ставил вопрос, не являются ли эти приступы результатом обратного воздействия каких-то быстрых волн, возбуждаемых в среде световыми корпускулами. Идея приступов навязывалась также *периодичностью световых процессов*, подмеченной Ньютоном при исследовании интерференционного явления «ньютоновых колец». Эту периодичность Ньютон также пытался объяснить, дополнив корпускулярную теорию волновыми представлениями.

7. Физики XVIII и начала XIX веков, забыв о колебаниях и сомнениях Ньютона, приняли корпускулярную теорию света. Сторонники волновой теории насчитывались единицами. Правда, среди них были Эйлер (1707–1783), Ломоносов (1711–1765) и Франклин (1706–1790), выдвигавшие возражения против корпускулярной и приводившие аргументы в пользу волновой теории света. Эйлер, например, утверждал, что если бы корпускулярная теория была справедлива, то из-за излучения света масса Солнца заметно уменьшалась бы, а это сказалось бы на движении планет. Франклин указывал, что по корпускулярной теории свет должен был бы оказывать давление на освещаемые тела, тогда как все попытки обнаружить это давление оканчивались безрезультатно. Слабость этих возражений состояла в том, что они не содержали количественных оценок ожидаемых эффектов. Впрочем, согласно современным представлениям о взаимосвязи между массой и энергией, потеря массы из-за излучения, а также величина светового давления совершенно не зависят от того, является ли свет потоком частиц или волновым процессом. Эти и аналогичные возражения, как показало последующее развитие физики, не могли решить спор между корпускулярной и волновой теориями света. Нужны были новые экспериментальные факты.

8. Начиная с XIX века, положение стало складываться в пользу волновой теории благодаря работам Юнга (1773–1829) и в особенности Френеля (1788–1827), систематически исследовавших явления интерференции и дифракции света. На основе волновых представлений была создана стройная теория этих явлений, выводы и предсказания которой полностью согласовывались с экспериментом. Объяснение прямолинейного распространения света содержалось в этой теории как частный случай. Были открыты и исследованы новые оптические явления: поляризация света при отражении (Малюс, 1808) и преломлении (Малюс и Био, 1811), угол полной поляризации (Брюстер, 1815), интерференция поляризованных лучей (Френель и Араго, 1816), количественные законы и теория отражения и преломления света (Френель, 1821), двойное преломление сжатым стеклом (Брюстер, 1815), двуосные кристаллы (Брюстер, 1815), законы и теория распространения света в двуосных кристаллах (Френель, 1821), вращение плоскости поляризации в кварце (Араго, 1811) и жидкостях (Био, 1815; оба явления исследовались далее Био, Брюстером и др.). Юнг (1807) измерил на опыте длину световой волны. Оказалось, что волны красного света длиннее, чем синего и фиолетового. Тем самым в волновой теории было дано экспериментально обоснованное объяснение цветов света, которое связывало это явление с длиной световой волны. (Такое объяснение предлагалось еще Эйлером, но он не мог указать, длина каких волн больше — красных или синих.) Юнг (1817) высказал также мысль о *поперечности световых волн*. К такому же заключению независимо от него пришел Френель (1821) и обосновал это заключение путем исследования поляризации света и интерференции поляризованных лучей. Все эти факты и в особенности явления интерференции и дифракции света находили непринужденное объяснение в рамках волновой теории света. Корпускулярная теория не могла противопоставить ничего эквивалентного и к началу 30-х годов XIX века была оставлена.

Смертельный удар корпускулярной теории в ее ньютоновской форме был нанесен в 1850 г. К этому времени Физо (1819–1896) и Фуко (1819–1868) впервые измерили скорость света лабораторными методами. Как мы указывали (см. пп. 2 и 5), по корпускулярной теории скорость света в воде больше, а по волновой теории меньше, чем в вакууме. В 1850 г. Фуко и независимо от него Физо и Бреге сравнили обе скорости. Опыт оказался в согласии с волновой и в противоречии с корпускулярной теориями света. Физики XIX века восприняли это как решающий опыт, окончательно доказавший неправильность корпускулярной теории света.

9. Волновая теория, конечно, не могла считаться полной, пока не была установлена *природа световых колебаний*, или колебаний мирового эфира, как говорили физики девятнадцатого (и отчасти первой четверти двадцатого) века. Они не сомневались, что эфир подчиняется обычным законам механики Ньютона и к нему применимы такие понятия, как плотность, упругость, пространственное перемещение, скорость, ускорение и пр. Они пытались вывести строение и свойства эфира из наблюдаемых явлений и экспериментально установленных законов оптики. Поперечность световых волн заставила приписать мировому эфиру свойства твердой среды. Это породило ряд трудностей, в частности в вопросе об отражении и преломлении света (подробнее см. § 63). Нет необходимости останавливаться на этих трудностях и попытках их преодоления в теории эфира. Все это уже давно потеряло актуальность и сохранило лишь исторический интерес.

В 60-х годах XIX века Максвеллом были установлены общие законы электромагнитного поля, которые привели его к заключению, что *свет — это электромагнитные волны* (см. т. III, гл. IV). Подтверждением такой точки зрения в то время были открытие Фарадеем в 1846 г. вращения плоскости поляризации света в магнитном поле и совпадение скорости света в вакууме с электродинамической постоянной, установленное на опыте в 1856 г. Вебером и Кольраушем (см. т. III, § 51 и 83). После известных опытов Герца (1887–1888 гг., см. т. III, § 142) электромагнитная природа света быстро получила признание. Уже в первом десятилетии XX века она из гипотезы превратилась в твердо установленный факт. Световые колебания были отождествлены с колебаниями электромагнитного поля. Оптика превратилась в раздел учения об электрических и магнитных явлениях.

Электромагнитная теория света устранила трудности, с которыми столкнулась теория упругого твердого эфира. Однако физики XIX века считали, что она дала не настоящее, а только *символическое* решение вопроса о природе света. Они смотрели на нее как на формальную схему, уравнения которой правильно передают количественные соотношения между различными величинами и явлениями, но символы, входящие в эти уравнения, еще не получили отчетливого физического истолкования. Считалось, что уравнения Максвелла должны составлять математический остов будущей более полной физической теории электромагнитных и световых явлений, но сами по себе они такой физической теории еще не составляют; последняя станет возможной лишь после того, как будут найдены механические свойства эфира.

Однако эти надежды не оправдались. Световой (или, лучше, электромагнитный) эфир упрямо отказывался обнаружить свои «механические свойства». В частности, потерпели неудачу настойчивые попытки обнаружить движение Земли относительно эфира, начатые Майкельсоном (1852–1931) в 1881 г. При изучении оптических и электродинамических явлений в движущихся средах обнаружились и другие расхождения теории эфира с опытом. Это привело Эйнштейна (1879–1955) в 1905 г. к теории относительности. Гипотеза механического эфира была оставлена[1]). Пошатнулась вера в «механическую картину мира», к построению которой так настойчиво стремились физики

---

[1]) Необходимо заметить, однако, что в вакууме нет обычного вещества, как оно понимается в химии. Но вакуум не есть пустота в буквальном смысле этого слова. Его заполняют физические поля (гравитационное, электромагнитное, ядерное и пр.). Они, наряду с обычным химическим веществом, являются различными формами материи. В вакууме могут происходить различные физические процессы. Примером может служить *поляризация вакуума*, т.е. рождение пар электрон–позитрон в сильных электрических полях. Можно было бы не возражать по существу против употребления термина «эфир» в смысле носителя этих физических свойств «пустого» пространства. Возражение относится к представлению об эфире как о жидкой, твердой, упругой или какой-либо другой среде, наделенной механическими свойствами. Однако в современной физике предпочитают не пользоваться термином «эфир» в указанном смысле, а употребляют термин «вакуум».

XIX века. И все последующее развитие науки привело физиков к убеждению в невозможности сведения всех явлений природы к механике. Утверждать противоположное — значит предъявлять к природе необоснованные требования. Поэтому, когда в современной волновой теории говорят, что свет — это колебания электромагнитного поля, то на это уже не смотрят как на формальное утверждение, а считают, что сами эти колебания не сводятся к чему-то «более простому и наглядному».

10. Важным этапом в развитии максвелловской электродинамики было введение в нее *атомистических представлений*, что было систематически проведено в *электронной теории* Г.А. Лорентца (1853–1928). В теории Максвелла вещество характеризовалось *феноменологическими постоянными* — диэлектрической и магнитной проницаемостями и удельной электрической проводимостью. В электронной теории эти макроскопические постоянные получили истолкование с атомистической точки зрения. В результате этого не только получили принципиальное, хотя и недостаточное, объяснение многие известные электродинамические и оптические явления (например, *дисперсия* и *абсорбция света*), но были открыты и объяснены и другие явления (*явления Керра*, *Зеемана*, *Фарадея*, *Коттона–Мутона*, *молекулярное рассеяние света* и т.д.).

Однако классическая физика и, в частности, электронная теория оказались недостаточными для истолкования явлений *атомного масштаба*. Потребовалось введение *квантовых представлений*. Необходимость и плодотворность последних обнаружилась ранее всего при изучении проблемы распределения энергии в спектре черного излучения, т.е. температурного излучения абсолютно черного тела. Применение к этой проблеме принципов классической физики приводило к глубоким противоречиям с опытом. Планк (1858–1947) в конце 1900 г. получил согласующуюся с опытом формулу для распределения энергии в спектре черного излучения. При этом он ввел чуждое классической физике представление, что излучение и поглощение света осуществляется не непрерывно, а *конечными порциями*, или *квантами энергии*, причем величина кванта определяется выражением (1.1). Для решения проблемы черного излучения Планку достаточно было принять, что этот квантовый характер излучения и поглощения света относится к *статистическим процессам*. Через пять лет Эйнштейн показал, что его необходимо распространить и на *элементарные процессы*. Согласно Эйнштейну, не только излучение и поглощение, но и *распространение света* в пространстве происходят конечными порциями — квантами света, обладающими определенной энергией и определенным импульсом. Так возродилось представление о частицах света, названных позднее *фотонами*.

Гипотеза фотонов позволила прежде всего объяснить загадочные закономерности в явлениях *фотоэффекта*, совершенно непонятные с точки зрения классической волновой теории света. Существование импульса у фотонов было доказано открытием в 1923 г. *эффекта Комптона* — изменения длины волны при рассеянии рентгеновского излучения. Гипотеза световых квантов позволила понять *химические действия света* и их закономерности. Квантовый характер излучения и поглощения света был использован Бором (1885–1962) для объяснения *спектральных закономерностей.*

Но как согласовать корпускулярные представления о свете с результатами опытов Фуко и Физо (см. п. 8)? Эти опыты вне всякого сомнения опровергают корпускулярную теорию света в ее ньютоновской форме. Приходится поэтому признать, что *к световым корпускулам классические представления о движении неприменимы.* Интерференция и дифракция света доказывают, что в этих явлениях свет ведет себя как *волны.* Фотоэффект, комптоновское рассеяние рентгеновских лучей и пр. с неменьшей убедительностью доказывают, что здесь свет действует как *частицы.* Вообще, *явления распространения света правильно описываются в рамках волновых теорий, а для описания взаимодействия света и вещества необходимы корпускулярные представления.* Этот «*дуализм волн и частиц*» надо рассматривать как экспериментальный факт, и поэтому полная теория света должна быть не корпускулярной и не волновой, а *корпускулярно-волновой.*

Такое парадоксальное положение еще более расширилось после того, как Дэвиссоном (1881–1958) и Джермером (1896–1971) в 1927 г. была открыта *дифракция электронов.* Оказалось, что *волновые свойства присущи и частицам обычного вещества,* — идея, развивавшаяся французским физиком де Бройлем (р. 1892) за несколько лет до открытия дифракции электронов. Развитие квантовой механики позволило частично объяснить возникшее парадоксальное положение ценой отказа от основного положения классической физики — *принципа причинности в форме детерминизма.* А исследования в области *физики высоких энергий* (иначе называемой *физикой элементарных частиц*) показали, что если энергия частиц превосходит их энергию покоя, то частицы могут рождаться, исчезать или превращаться друг в друга. В этом отношении они ведут себя подобно фотонам, которые могут излучаться или поглощаться. В квантовой электродинамике фотоны рассматриваются как *кванты электромагнитного поля.* Поэтому в физике высоких энергий целесообразно говорить об *электронно-позитронном, мезонном, нуклонном* и прочих полях, квантами которых являются электроны, позитроны, мезоны, протоны, нейтроны и т.д.

Таким образом, вопрос о природе света стал частью более общей проблемы строения вещества.

§ 4. Искривление световых лучей в неоднородных средах

1. В неоднородных средах представление о распространении света вдоль лучей сохраняется, но сами лучи становятся *криволинейными.* Действительно, рассмотрим среду, состоящую из плоскопараллельных слоев с постоянными показателями преломления (рис. 14), меняющимися скачкообразно от слоя к слою. Световой луч, преломляясь на границах слоев, примет форму ломаной линии. Будем неограниченно увеличивать число слоев, устремляя к нулю их толщины и скачки показателя преломления. Тогда в пределе показатель преломления среды станет меняться в пространстве непрерывно, а луч перейдет в кривую с непрерывно изменяющейся касательной. Допустим теперь, что показатель преломления меняется в пространстве как угодно, но не слишком резко (см. следующий параграф). Проведем в среде поверхности равного показателя преломления. В пределах каждого достаточно малого объема эти поверхности можно считать плоскими, а среду плоскослоистой, к которой применимо рассуждение, приведенное выше.

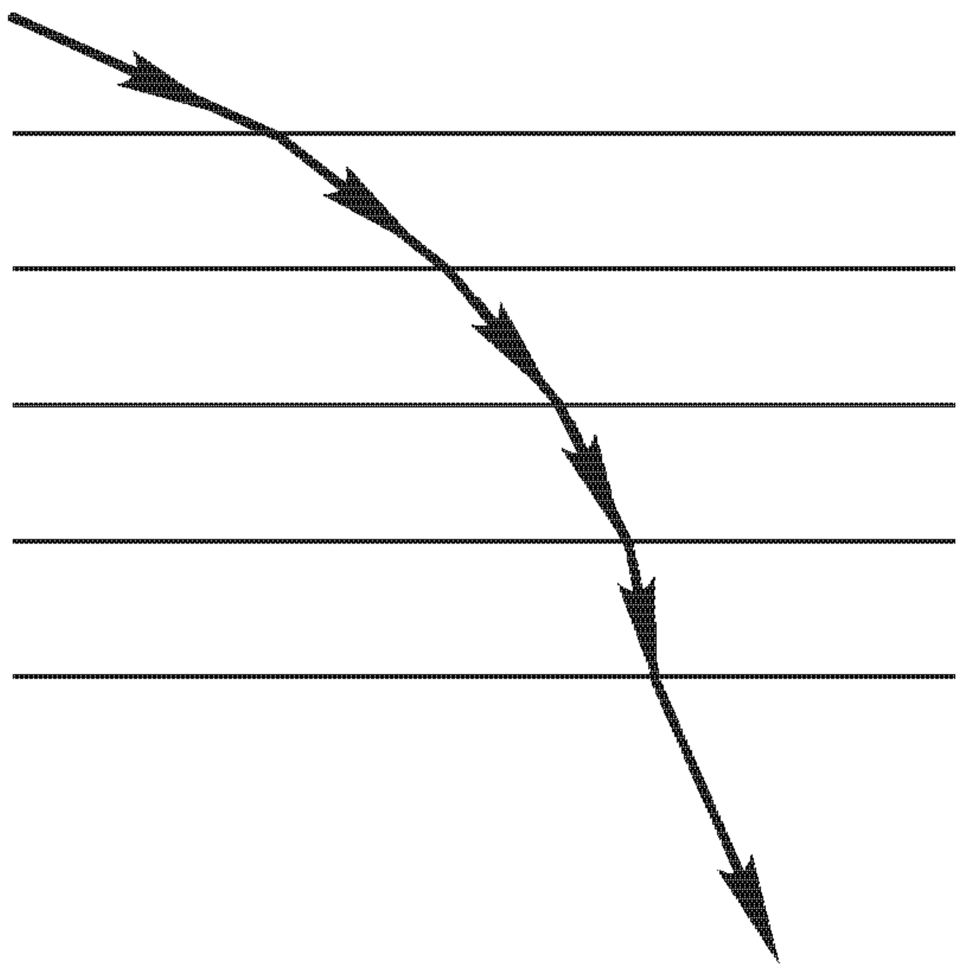


Рис. 14

Из изложенного следует, что геометрическую форму луча можно однозначно определить из закона Снеллиуса путем предельного перехода. Но закон Спеллиуса получается не только в волновой, но и в корпускулярной теории Ньютона. Поэтому при определении формы светового луча можно рассуждать так, как если бы свет состоял из ньютоновых корпускул, а показатель преломления n определялся формулой (3.3). Поскольку преломление определяется относительным показателем преломления, абсолютный показатель можно заменить величиной, ему пропорциональной. Ради краткости можно просто положить $n = v$. Скорость корпускулы v однозначно определяется уравнением сохранения энергии, а потому v можно рассматривать как известную функцию координат. Таким образом, *траектория корпускулы в потенциальном поле сил геометрически совпадает с лучом света в среде*, показатель преломления n которой численно равен v. Для этого, конечно, необходимо, чтобы исходные направления этих двух кривых были одинаковы. Эта формальная аналогия между движением частицы и распространением светового луча позволяет перенести результаты, полученные в световой оптике, в *электронную микроскопию*, где роль световых лучей выполняют электроны, движущиеся в потенциальных электрических полях.

Воспользуемся отмеченной аналогией для вычисления радиуса кривизны R светового луча. Нормальное ускорение корпускулы определяется формулой

$$\frac{mv^2}{R} = F_N = -\frac{\partial U}{\partial N},$$

где F_N — составляющая действующей силы $\mathbf{F}$ вдоль единичного вектора главной нормали $\mathbf{N}$, а U — потенциальная энергия корпускулы. Дифференцируя вдоль нормали $\mathbf{N}$ уравнение энергии $^1/_2mv^2 + U = \text{const}$, получаем $F_N = mv\,\partial v/\partial N$, а потому

$$\frac{1}{R} = \frac{1}{v}\frac{\partial v}{\partial N}.$$

Заменяя v на n, находим выражение для кривизны луча:

$$\frac{1}{R} = \frac{1}{n}\frac{\partial n}{\partial N} = \frac{\partial}{\partial N}(\ln n). \qquad (4.1)$$

При этом ускорение корпускулы, а с ним и сила $\mathbf{F}$ не имеют составляющей вдоль бинормали $\mathbf{b}$ к траектории, т.е. $F_b = -\partial U/\partial b = mv\,\partial v/\partial b = 0$. Отсюда $\partial v/\partial b = 0$, а потому также $\partial n/\partial b = 0$. Значит, вектор $\operatorname{grad} n$ лежит в соприкасающейся плоскости светового луча. Поэтому из всех направлений, перпендикулярных к лучу, направление главной нормали $\mathbf{N}$ характеризуется самым быстрым изменением показателя преломления среды. Это значит, что *в неоднородной среде луч изгибается в сторону наиболее быстрого изменения показателя преломления*. Если среда однородна ($n = \text{const}$), то кривизна $1/R$ обращается в нуль, т.е. световые лучи прямолинейны.

2. Плотность земной атмосферы, а с ней и показатель преломления убывают с высотой. Этим объясняется ряд явлений, связанных с искривлением световых лучей. К ним относится, например, *астрономическая рефракция*, т.е. кажущееся поднятие небесного светила из-за искривления световых лучей в земной атмосфере. Если светило стоит высоко над горизонтом, то при расчете этого явления мы не сделаем заметной ошибки, считая поверхность Земли плоской. Но так поступать нельзя, когда светило находится вблизи горизонта. В этом случае надо учитывать сферичность земной поверхности.

Пренебрежем малыми боковыми градиентами показателя преломления воздуха n и будем считать, что n зависит только от высоты над земной поверхностью или, что то же самое, от расстояния r от центра земного шара O (рис. 15). Световой луч AM от небесного светила будет лежать в вертикальной плоскости, проходящей через это светило и глаз наблюдателя. На основании (4.1)

$$\frac{1}{R} = \frac{d\alpha}{ds} = \frac{\partial}{\partial N}(\ln n) = \frac{d(\ln n)}{dr}\frac{\partial r}{\partial N} = -\frac{d(\ln n)}{dr}\sin\beta,$$

где α — так называемое *зенитное расстояние*, т.е. переменный угол между вертикалью места MZ и касательной к лучу, а s — длина луча, отсчитываемая от глаза наблюдателя M. Как видно из рис. 15,

$$\frac{d\alpha}{ds} = \frac{d\alpha}{dr}\frac{dr}{ds} = \frac{d\alpha}{dr}\cos\beta,$$

а потому

$$\frac{1}{\operatorname{tg}\beta}\frac{d\alpha}{dr} = -\frac{d}{dr}(\ln n). \tag{4.2}$$

Далее, $r\,d\gamma = ds\sin\beta$, откуда

$$\frac{1}{\operatorname{tg}\beta}\frac{d\gamma}{dr} = \frac{1}{r}\frac{ds}{dr}\cos\beta = \frac{1}{r}.$$

Вычитая это равенство из (4.2) и принимая во внимание, что $\alpha - \gamma \approx \beta$, получим

$$\frac{d\beta}{\operatorname{tg}\beta} = -d\ln n - \frac{dr}{r} = -d\ln(nr).$$

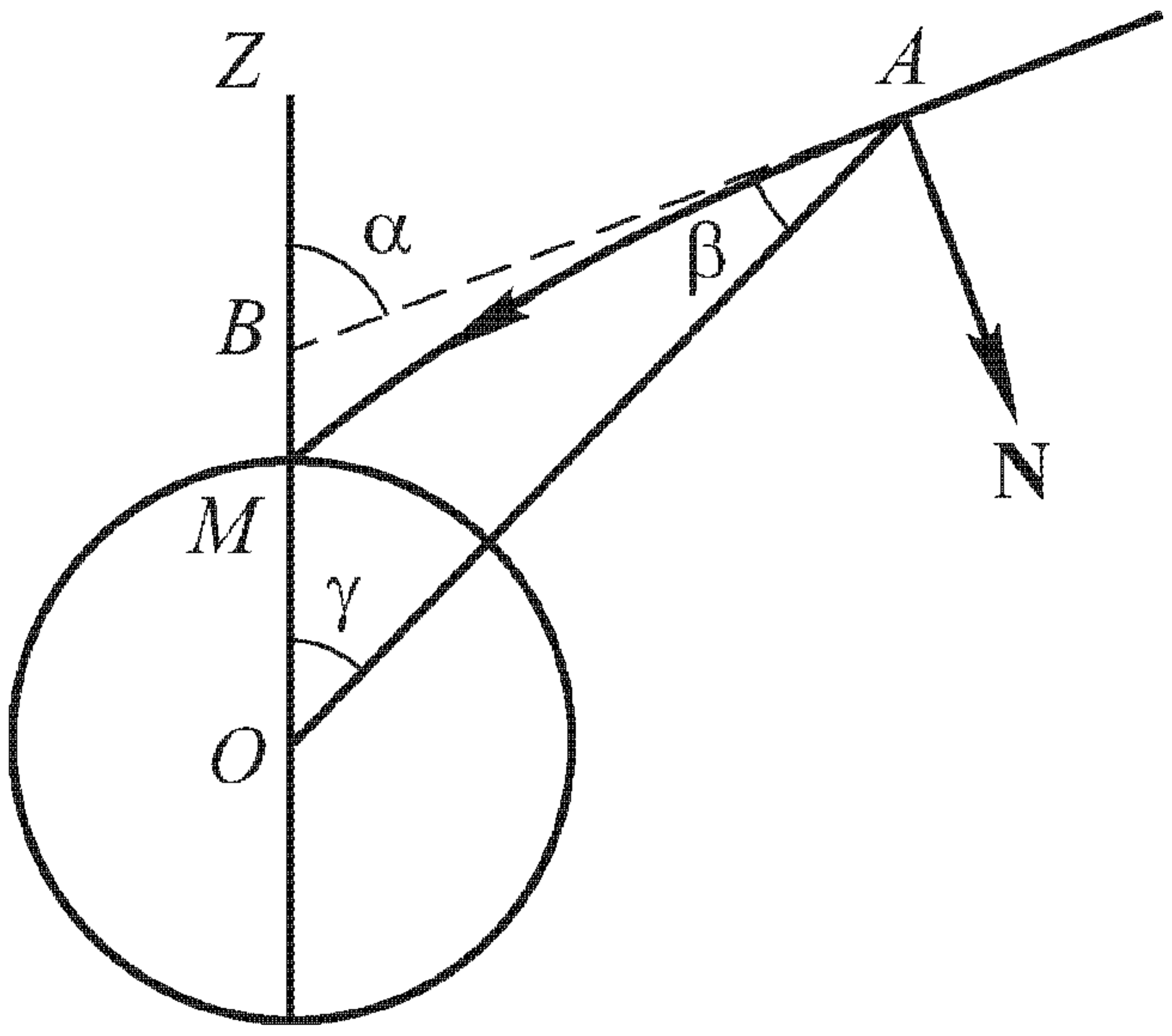


Рис. 15

Интегрирование этого уравнения дает

$$nr\sin\beta = n_0 r_0 \sin\alpha_0, \tag{4.3}$$

где нулем обозначены величины n, r, α в точке M (в которой $\alpha \equiv \beta$). Вычислив отсюда $\operatorname{tg}\beta$ и подставив его значение в (4.2), найдем

$$\alpha_\infty - \alpha_0 = -n_0 r_0 \sin\alpha_0 \int\limits_{r_0}^{\infty} \frac{d\ln n}{dr}\frac{dr}{\sqrt{n^2r^2 - n_0^2r_0^2\sin^2\alpha_0}}. \tag{4.4}$$

Здесь α_0 — видимое зенитное расстояние светила в точке M, а α_∞ — угол между асимптотой луча и вертикалью в той же точке. Разность $\alpha_\infty - \alpha_0$ называется *рефракцией*. Для ее вычисления и служит формула (4.4). Зная зависимость плотности воздуха ρ от высоты над земной поверхностью, находят n по формуле $(n-1)/\rho = \mathrm{const}$, после чего производят численное интегрирование в формуле (4.4).

Если светило находится не слишком близко от горизонта, то кривизной земной поверхности можно пренебречь. Тогда рефракция не зависит от характера изменения показателя преломления с высотой и может быть вычислена по формуле

$$\sin\alpha_\infty = n_0\sin\alpha_0 \tag{4.5}$$

(см. задачу 3 к § 2). Ошибка, даваемая этой формулой, быстро убывает с уменьшением α. Уже при $\alpha_0 = 80°$ она не превосходит 3 %. При $\alpha_0 < 75°$ формула (4.5) дает достаточную точность.

Если светило находится на горизонте ($\alpha = 90\,°$), то средняя рефракция при 10 °C и 760 мм рт. ст. составляет $35'24''$. Она быстро уменьшается по мере поднятия светила над горизонтом. Уже при $\alpha_0 = 89°$ средняя рефракция уменьшается до $24'37''$. Этим объясняется сплюснутая форма Солнца при восходе и заходе.

С рефракцией связано некоторое удлинение дня. Для средних широт оно составляет в среднем 3–4 минуты.

Рефракция зависит от длины волны. Поэтому при заходе Солнца сначала должны исчезать красные и желтые лучи, а оставшийся сегмент солнечного диска должен окрашиваться на одну-две секунды в зеленый или даже синий цвет. При восходе Солнца, наоборот, должна сначала появляться кратковременная зеленая вспышка. Это явление «зеленого луча» наблюдается на море, да и то крайне редко, так как необходимы исключительно спокойная атмосфера и ясная погода.

3. Иногда вблизи земной поверхности из-за сильного нагревания или охлаждения возникают большие градиенты показателя преломления воздуха. Тогда шарообразность Земли можно не учитывать. Если градиент n направлен вертикально, то можно воспользоваться формулой (4.3), полагая в ней $\beta = \alpha$, $r = r_0$. В результате получится

$$n \sin\alpha = n_0 \sin\alpha_0 = \text{const}. \tag{4.6}$$

Допустим ради определенности, что световой луч распространяется вверх под малым углом к горизонту, а показатель преломления $n = n(z)$ убывает с высотой z. Может случиться, что на некоторой высоте угол α обратится в 90°. Тогда касательная к лучу станет горизонтальной. Высота $z = h$, на которой это может произойти, определится из соотношения

$$n(h) = n_0 \sin\alpha_0.$$

Так как показатель преломления возрастает вниз, то, достигнув высоты $z = h$, луч в дальнейшем должен загнуться книзу. На указанной высоте происходит нечто аналогичное полному отражению (рис. 16). Такое явление может возникнуть при сильно аномальном распределении плотности воздуха по высоте и является причиной различного рода *миражей*, наблюдаемых в атмосфере.

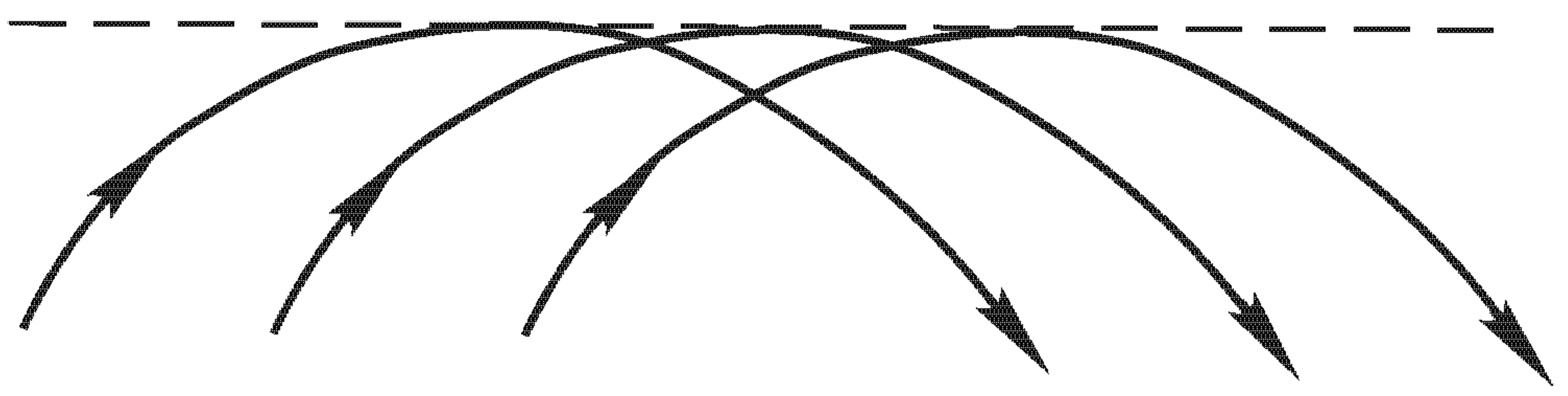


Рис. 16

Обычно наблюдается *верхний* или *нижний мираж*. При верхнем мираже, помимо самих предметов, видны их изображения, расположенные сверху; при нижнем мираже изображение получается ниже самого предмета. Нижний мираж наблюдается в пустынях и в степях в теплое время года, когда прилегающий к земной поверхности слой воздуха сильно нагрет, а его плотность и показатель преломления быстро возрастают с высотой. Из каждой точки предмета в глаз наблюдателя всегда попадают прямые лучи, т.е. лучи, не испытавшие полного отражения в воздухе; им соответствует обычное — прямое — изображение предмета. Но при больших градиентах показателя преломления могут также попасть лучи, испытавшие полное отражение. Они дают обратное изображение предмета, как в зеркале (рис. 17). При этом лучи, выходящие из различных точек предмета, претерпевают полное

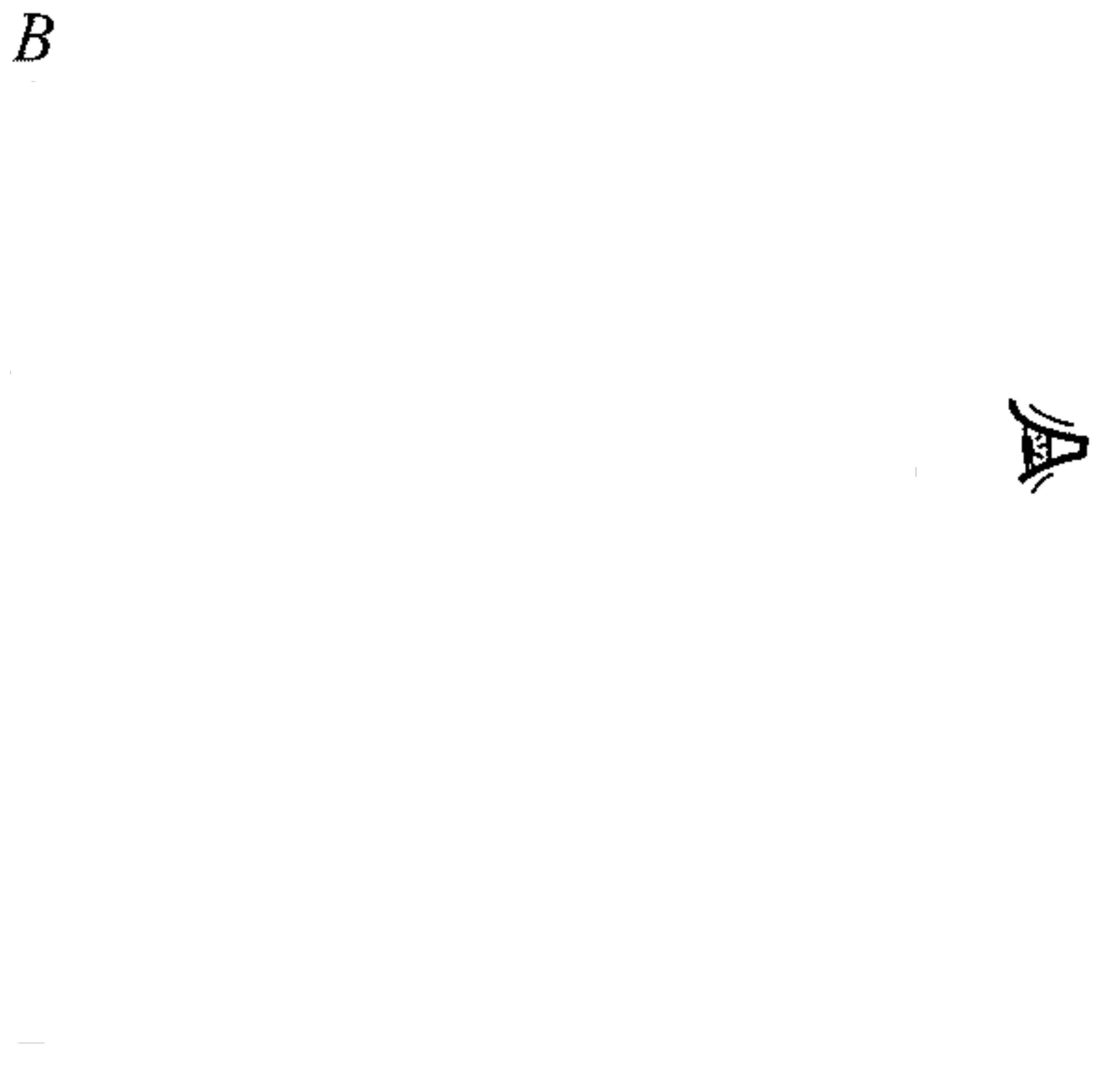


Рис. 17

отражение на несколько разных высотах; точка M, в которой отражается луч, выходящий из A, лежит несколько выше точки N, в которой отражается луч, выходящий из B. Наблюдатель видит два изображения: прямое и обратное. Создается иллюзия водной поверхности, в которой, как в зеркале, видно изображение неба.

Аналогично объясняется и верхний мираж. Он наблюдается зимой в холодных странах, когда вблизи земной поверхности образуется холодный слой воздуха, в котором показатель преломления быстро убывает с высотой. В горах, хотя и очень редко, наблюдается *боковой мираж*, связанный с изменением показателя преломления воздуха в боковом направлении. Вообще, в зависимости от характера распределения показателя преломления воздуха, мираж проявляется в весьма разнообразных и часто причудливых формах.

ЗАДАЧИ

1. Показать, что в пренебрежении кривизной земной поверхности справедлива формула

$$\alpha_\infty - \alpha_0 = (n_0 - 1)\operatorname{tg}\alpha_0. \tag{4.7}$$

2. Определить, при каких градиентах температуры T воздуха возможен нижний мираж.

Решение. Показатель преломления, а с ним и плотность воздуха ρ должны увеличиваться с высотой h (т.е. $d\rho/dh > 0$). Воспользуемся уравнением состояния $\rho = \mu P/(RT)$, где P — давление, μ — относительная молекулярная масса воздуха, а R — универсальная газовая постоянная. Из него находим

$$\frac{1}{\rho}\frac{d\rho}{dh} = \frac{1}{P}\frac{dP}{dh} - \frac{1}{T}\frac{dT}{dh}.$$

При механическом равновесии воздуха $dP/dh = -\rho g$, где g — ускорение свободного падения. В результате получаем

$$\frac{dT}{dh} < \frac{\mu g}{R}. \tag{4.8}$$

Используя соотношение $c_P - c_v = R/\mu$, этому условию можно придать вид

$$\frac{dT}{dh} < -\frac{g}{c_P - c_v} \approx -0{,}025\ \text{К/м}, \tag{4.9}$$

где c_P и c_v — удельные теплоемкости воздуха при постоянном давлении и постоянном объеме. Такое распределение температур *конвективно неустойчиво*, так как для конвективной устойчивости необходимо

$$\frac{dT}{dh} \geqslant -\frac{g}{c_P} \approx -0{,}01\ \text{К/м}$$

(см. т. II, § 121).

§ 5. Плоские электромагнитные волны

1. С электромагнитными волнами мы встречались уже в томе III нашего курса (гл. X). Выведем еще раз основные результаты, полученные в § 139 указанного тома, предполагая здесь, что волны — *монохроматические*. Будем представлять их в комплексной форме. В связи с этим полезно еще раз обратиться к § 126 того же тома, где введены комплексные обозначения и изложены правила оперирования с ними. По причинам, которые выяснятся в ходе изложения, монохроматические волны и их комплексная форма имеют важнейшее значение в учении о волнах вообще, а в особенности в оптике.

Допустим, что волны распространяются в однородном незаряженном диэлектрике. Применим к ним фундаментальные уравнения Максвелла

$$\operatorname{rot}\mathbf{H} = \frac{1}{c}\frac{\partial\mathbf{D}}{\partial t}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} \tag{5.1}$$

и материальные уравнения

$$\mathbf{D} = \varepsilon\mathbf{E}, \quad \mathbf{B} = \mu\mathbf{H}. \tag{5.2}$$

Пусть волна — плоская и монохроматическая. Запишем ее в комплексном виде

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{kr})}, \quad \mathbf{H} = \mathbf{H}_0 e^{i(\omega t - \mathbf{kr})}, \tag{5.3}$$

где ω — круговая частота, $\mathbf{k}$ — волновой вектор, а амплитуды $\mathbf{E}_0$ и $\mathbf{H}_0$ постоянны. Дифференцируя по времени, получаем $\partial\mathbf{E}/\partial t = i\omega\mathbf{E}$, т.е. операция дифференцирования в этом случае сводится к умножению на $i\omega$. Аналогично, дифференцирование по координатам x, y, z сводится к умножению на $-ik_x$, $-ik_y$, $-ik_z$. Заметив это и обозначая координатные орты через $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$, получаем

$$\operatorname{rot}\mathbf{H} = \begin{vmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ H_x & H_y & H_z \end{vmatrix} = -i \begin{vmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \\ k_x & k_y & k_z \\ H_x & H_y & H_z \end{vmatrix} = -i[\mathbf{kH}]$$

и аналогично для $\operatorname{rot}\mathbf{E}$. В результате уравнения (5.1) перейдут в

$$[\mathbf{kH}] = -\frac{\omega}{c}\mathbf{D}, \quad [\mathbf{kE}] = \frac{\omega}{c}\mathbf{B}. \tag{5.4}$$

Введем единичный вектор $\mathbf{N}$ нормали к фронту волны и скорость распространения последнего в направлении этой нормали — так называемую *нормальную скорость* v. Тогда

$$\mathbf{k} = \frac{\omega}{v}\mathbf{N}, \tag{5.5}$$

и предыдущие соотношения перейдут в

$$\mathbf{D} = -\frac{c}{v}[\mathbf{NH}], \quad \mathbf{B} = \frac{c}{v}[\mathbf{NE}]. \tag{5.6}$$

Отсюда видно, что векторы $\mathbf{E}$, $\mathbf{H}$, $\mathbf{v}$ в плоской электромагнитной волне взаимно перпендикулярны. Их взаимное расположение представлено на рис. 18. Перпендикулярность векторов **E** и **H** к вектору **v**, или, что то же, к направлению распространения волны, означает, что *электромагнитные волны поперечны.* Таким образом, проблема поперечности световых волн, с которой не могли справиться теории механического эфира, совсем не возникает в электромагнитной теории света.

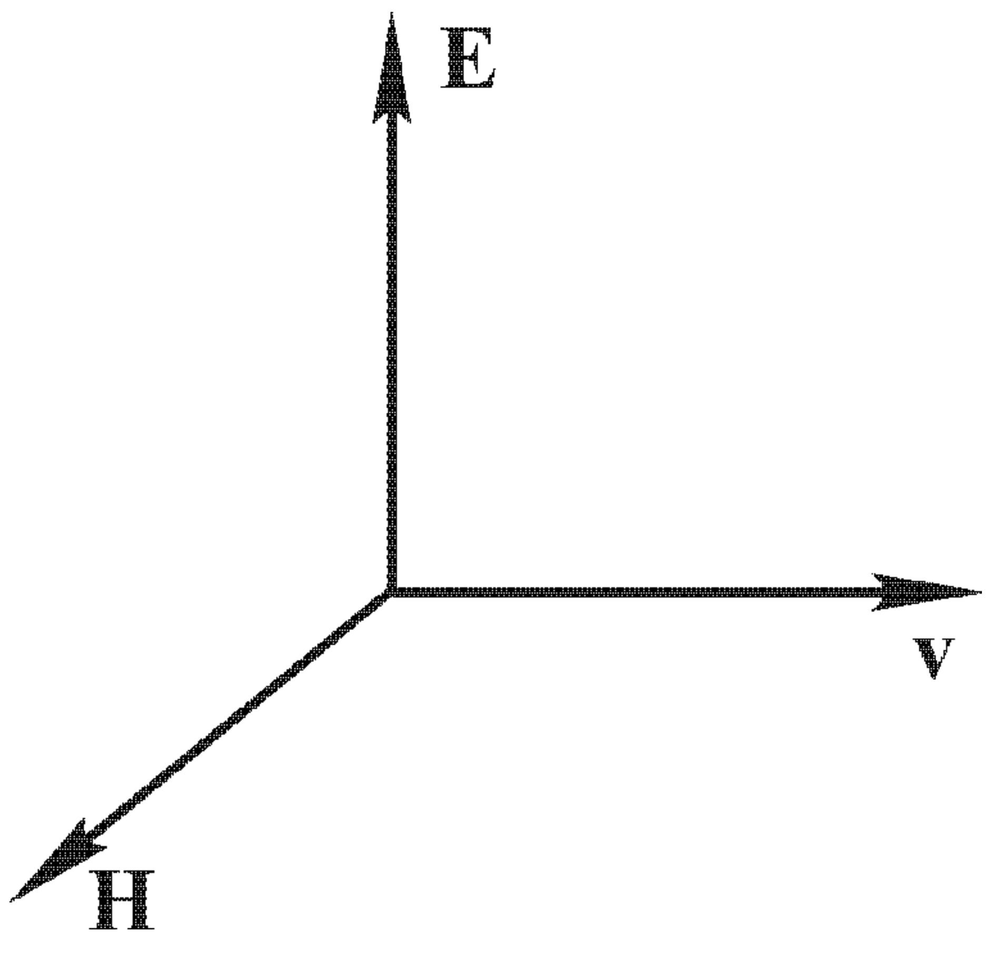


Рис. 18

Следует отметить, что между взаимно перпендикулярными составляющими, на которые можно разложить комплексные амплитуды $\mathbf{E}_0$ и $\mathbf{H}_0$, может существовать какая-то разность фаз φ. Если она равна нулю или π, то вектор **E** во всех точках пространства и во все моменты времени будет находиться в одной и той же плоскости, проходящей через направление нормали **N**, а вектор **H** — в плоскости, к ней перпендикулярной. Тогда волну называют *линейно-* или *плоскополяризованной.* Плоскость $(\mathbf{E},\mathbf{N})$, в которой лежат векторы **E** и **N**, называется *плоскостью колебаний* или *плоскостью поляризации волны*[1]. Если φ не равна нулю или π, то возникает так называемая *эллиптическая поляризация.* Исследованием ее мы займемся в главах V, VI и VII.

Из уравнений (5.1) можно определить и скорость электромагнитной волны v. С этой целью запишем эти уравнения в скалярной форме:

$$D = \frac{c}{v}H, \quad B = \frac{c}{v}E, \tag{5.7}$$

или

$$\varepsilon E = \frac{c}{v}H, \quad \mu H = \frac{c}{v}E. \tag{5.8}$$

Отсюда после почленного перемножения и сокращения на EH получаем для v и показателя преломления $n = c/v$ следующие выражения:

$$v = c/\sqrt{\varepsilon\mu}, \tag{5.9}$$

$$n = \sqrt{\varepsilon\mu}. \tag{5.10}$$

Последнее соотношение называется *законом Максвелла.* Для немагнитных сред ($\mu = 1$) оно переходит в

$$n = \sqrt{\varepsilon}. \tag{5.11}$$

---

[1] Раньше плоскостью поляризации называли плоскость $(\mathbf{H},\mathbf{N})$, перпендикулярную к $(\mathbf{E},\mathbf{N})$. Но этот термин в таком понимании выходит из употребления.

В вакууме $v = c$, т.е. v совпадает с *электродинамической постоянной* c. Тем самым раскрывается глубокий физический смысл открытия В. Вебера (1804–1891) и Р. Кольрауша (1809–1858), впервые измеривших эту постоянную в 1856 г. (см. т. III, § 51, 83).

2. Обратимся теперь к экспериментальной проверке соотношения (5.11). В табл. 1 сопоставлены экспериментально измеренные значения n и $\sqrt{\varepsilon}$ для ряда веществ (показатели преломления относятся к желтой линии натрия). Для газов, приведенных в этой таблице, закон Максвелла (5.11) хорошо согласуется с опытом. Для жидких углеводородов согласие хуже. Для воды и спиртов, а также для большинства других твердых и жидких тел наблюдаются резкие нарушения соотношения (5.11). Однако в этом нет ничего неожиданного. Дело в том, что значения ε, приведенные в табл. 1, относятся к *статическим электрическим полям*, а значения n — к *электромагнитным полям световых волн*, частоты которых порядка $5 \cdot 10^{14}$ Гц. Диэлектрическая проницаемость ε обусловлена *поляризацией диэлектрика*, т.е. смещением заряженных частиц внутри атомов и молекул под действием внешнего электрического поля. Для правильного сопоставления надо брать значения ε, измеренные в электрических полях *тех же частот*. Действительно, атомы и молекулы обладают собственными частотами, так что амплитуды (и фазы) вынужденных колебаний электронов и ядер, из которых они состоят, зависят от частоты внешнего электрического поля. Особенно сильную зависимость следует ожидать в тех случаях, когда частота внешнего поля близка к одной из собственных частот атомов или молекул (резонанс!). В результате возникает зависимость показателя преломления вещества от частоты световой волны — так называемая *дисперсия света*.

Таблица 1

Вещество	n	$\sqrt{\varepsilon}$	Вещество	n	$\sqrt{\varepsilon}$
Воздух	1,000292	1,000302	Толуол	1,499	1,549
Азот	1,000299	1,000307	Четыреххлористый углерод	1,461	1,523
Кислород	1,000270	1,000273			
Водород	1,000139	1,000139	Бензол	1,501	1,511
Углекислота	1,000499	1,000485	Сероуглерод	1,629	1,626
Гелий	1,000035	1,000037	Парафин	1,422	1,405
Окись углерода	1,000335	1,000346	Вода	1,33	9,00
Аммиак	1,000385	1,000385	Метиловый спирт	1,34	5,70
Закись азота	1,000507	1,000547	Этиловый спирт	1,36	5,10

Аналогично, намагниченность вещества (парамагнетизм и ферромагнетизм) возникает в результате *поворотов магнитных моментов атомов и молекул* под действием внешнего магнитного поля. Однако в полях столь высоких частот, которые лежат в оптической области спектра, атомы и молекулы не успевают поворачиваться за времена порядка периода световых колебаний. Диамагнитный же эффект, принципиально имеющий место во всех веществах, пренебрежимо мал. Поэтому в полях указанных частот намагничивание вещества практически не происходит. Вот почему в оптических явлениях, за редкими исключениями, магнитные свойства вещества не проявляются, и можно пользоваться формулой (5.11) вместо более общей формулы (5.10).

Таким образом, расхождения закона Максвелла (5.10) или (5.11) с опытом происходят за счет *нарушения материальных уравнений* (5.2). Такие уравнения справедливы не всегда, а только для *монохроматических полей*, причем ε *и* μ *являются функциями частоты электромагнитного поля*, различными для различных веществ. Чтобы отметить это обстоятельство, величины $\varepsilon(\omega)$ и $\mu(\omega)$ часто называют *динамическими диэлектрической и магнитной проницаемостями*, в отличие от *статических проницаемостей*, в которые они переходят при $\omega = 0$. Лишь в области сравнительно длинных электромагнитных волн (превышающих примерно 1 см) функции $\varepsilon(\omega)$ и $\mu(\omega)$ становятся постоянными для всех веществ. Поэтому в оптике электромагнитное поле приходится разлагать на монохроматические составляющие, что всегда возможно, согласно математической теореме Фурье (см. т. III, § 128). Предполагая, что выполняется принцип суперпозиции, эти монохроматические составляющие можно рассматривать *независимо друг от друга.* Таким путем можно исследовать распространение электромагнитных волн любого спектрального состава. Функции можно разлагать не только по синусам и косинусам, но и по бесконечному множеству других, «полных систем» функций. Однако выполнение материальных уравнений (5.2) для монохроматических полей, а также многие другие причины делают в оптике разложение полей на монохроматические составляющие *физически выделенным* среди множества других математически возможных разложений. Изложенные соображения, как и соображения, излагаемые в следующем пункте, имеют, конечно, общее значение, а не только для плоских электромагнитных волн.

3. Соблюдение материальных уравнений (5.2) (для монохроматических полей) предполагает, конечно, что величина ε не зависит от $\mathbf{E}$. Благодаря этому уравнения поля *линейны и однородны.* Однако это справедливо только в *слабых полях*, т.е. таких полях, напряженность которых весьма мала по сравнению с напряженностями внутриатомных и внутримолекулярных полей (10^7–10^8 В/см). В *сильных полях*, где это условие не соблюдается, уравнения поля в средах становятся *нелинейными*, что ведет к *нарушению принципа суперпозиции*, а следовательно, и закона независимости световых пучков. Оптику слабых электромагнитных полей (в указанном выше смысле) называют *линейной оптикой*, а оптику сильных полей — *нелинейной оптикой.*

До недавнего времени источники света, которыми располагала оптика, позволяли получать световые пучки лишь «слабой» интенсивности с максимальными электрическими полями 0,1–10 В/см (см. задачу к этому параграфу). Нелинейные эффекты (за редкими исключениями) в этих случаях совсем незаметны. Случай сильных электрических полей в оптике считался чисто умозрительным и долгое время не исследовался. Начиная с 1960 г. — года изобретения *оптических квантовых генераторов* (*лазеров*) — положение изменилось. Эти источники света позволяют получать световые волны, в которых электрические поля достигают 10^5–10^7 В/см. Такие поля уже не могут считаться пренебрежимо малыми по сравнению с внутриатомными и внутримолекулярными полями. В них наблюдаются качественно новые *«нелинейные явления»* и притом не как малые поправки к «линейным эффектам», а как *явления крупного масштаба*, получившие уже важные практические применения. О нелинейной оптике мы будем говорить в гл. XI. Во всех остальных главах излагается линейная оптика.

4. Для некоторых целей требуется более подробное исследование системы уравнений (5.4), которое мы и произведем. Запишем эту систему с учетом материальных уравнений (5.2):

$$[\mathbf{kH}] = -\frac{\omega\varepsilon}{c}\,\mathbf{E}, \qquad [\mathbf{kE}] = \frac{\omega\mu}{c}\,\mathbf{H}. \tag{5.12}$$

Подставляя **H** из второго уравнения в первое, получим

$$\frac{c}{\omega\mu}\,[\mathbf{k}[\mathbf{kE}]] = -\frac{\omega\varepsilon}{c}\,\mathbf{E}.$$

Так как векторы **k** и **E** взаимно перпендикулярны, то после раскрытия двойного векторного произведения и сокращения на **E** отсюда найдем

$$\mathbf{k}^2 = \frac{\omega^2}{c^2}\,\varepsilon\mu, \tag{5.13}$$

или, пренебрегая намагничиванием вещества,

$$\mathbf{k}^2 = \frac{\omega^2}{c^2}\,\varepsilon. \tag{5.14}$$

Таково условие совместности уравнений (5.12). Так как по самому определению волнового вектора $k = \omega/v$, то из (5.13) следует: $v = c/\sqrt{\varepsilon\mu}$, т.е. прежний результат (5.9).

Однако условиям (5.13) или (5.14) можно удовлетворить и в тех случаях, когда волновой вектор $\mathbf{k}$ *комплексный*, т.е. имеет вид

$$\mathbf{k} = \mathbf{k}' - i\mathbf{k}'', \tag{5.15}$$

где векторы $\mathbf{k}'$ и $\mathbf{k}''$ — вещественные. В этом случае волну называют *неоднородной*, в отличие от *однородной волны*, у которой волновой вектор $\mathbf{k}$ вещественный. Электрический вектор неоднородной волны можно представить в виде

$$\mathbf{E} = \mathbf{E}_0 e^{-\mathbf{k}''\mathbf{r}} e^{i(\omega t - \mathbf{k}'\mathbf{r})}. \tag{5.16}$$

Это выражение можно рассматривать как волну, амплитуда которой равна $\mathbf{E}_0 e^{-\mathbf{k}''\mathbf{r}}$. Она экспоненциально убывает в направлении вектора $\mathbf{k}''$. Для такой волны можно говорить о *поверхностях равных амплитуд* и *поверхностях равных фаз*. Поверхности равных амплитуд суть плоскости, перпендикулярные к вектору $\mathbf{k}''$. Поверхности равных фаз также плоскости, но перпендикулярные к вектору $\mathbf{k}'$. *В непоглощающей изотропной среде плоскости равных амплитуд и равных фаз взаимно перпендикулярны.* Для доказательства подставим выражение (5.15) в формулу (5.14). Отделяя вещественную часть от мнимой, найдем

$$\mathbf{k}'^2 - \mathbf{k}''^2 = \frac{\omega^2}{c^2}\varepsilon, \quad (\mathbf{k}'\mathbf{k}'') = 0. \tag{5.17}$$

Из второго соотношения видно, что векторы $\mathbf{k}'$ и $\mathbf{k}''$, а следовательно и плоскости равных амплитуд и фаз, взаимно перпендикулярны. Для поглощающих сред это утверждение, вообще говоря, не справедливо.

Чтобы получить более отчетливое представление о неоднородной волне, запишем ее в вещественной форме. Направим ось X вдоль вектора $\mathbf{k}'$, а ось Z — вдоль вектора $\mathbf{k}''$. Кроме того, положим $E_{0x} = A_x \exp(i\delta_x)$ и аналогично для E_{0y} и E_{0z}. Тогда неоднородную волну (5.16) в вещественной форме можно представить так:

$$\begin{aligned}
E_x &= A_x e^{-k''z} \cos(\omega t - k'x + \delta_x),\\
E_y &= A_y e^{-k''z} \cos(\omega t - k'x + \delta_y),\\
E_z &= A_z e^{-k''z} \cos(\omega t - k'x + \delta_z).
\end{aligned}$$

Плоскости равных фаз распространяются в направлении оси X со скоростью $v' = \omega/k'$. Она меньше скорости однородных волн $c/\sqrt{\varepsilon}$, поскольку $k' > \omega\sqrt{\varepsilon}/c$, как это следует из (5.17). Амплитуда убывает в направлении оси Z. Когда $z \to -\infty$, амплитуда неоднородной волны возрастает неограниченно. Поэтому плоские неоднородные волны в неограниченной среде существовать не могут. Но они могут при определенных условиях возникать вблизи границы среды. Так, например, такие волны возникают в оптически менее плотной среде при полном отражении света (см. § 66). Поле неоднородной волны заметно лишь в пограничном слое, толщина которого порядка длины волны. По этой причине неоднородные волны называют также *поверхностными волнами.*

ЗАДАЧА

Оценить напряженность поля солнечного излучения вблизи земной поверхности, если величина солнечной постоянной составляет около 2 кал$\cdot$см$^{-2}$ мин$^{-1}$ = 1,39$\cdot 10^6$ эрг$\cdot$см$^{-2}$с$^{-1}$. *Солнечной постоянной* называется количество энергии, попадающей от Солнца (при его среднем удалении от Земли) за единицу времени на единицу площади земной поверхности, перпендикулярной к излучению (при отсутствии абсорбции в атмосфере).

Решение. Волну, излучаемую Солнцем, у земной поверхности можно считать плоской. В такой волне $E = H$, поскольку для вакуума $\varepsilon = \mu = 1$. Плотность потока энергии $\frac{c}{4\pi}EH = \frac{c}{4\pi}E^2$. Приравнивая среднее значение этой величины значению солнечной постоянной, получим

$$\frac{c}{4\pi}\overline{E^2} = 1{,}39 \cdot 10^6, \text{ откуда } \overline{E^2} = 5{,}85 \cdot 10^{-4}, \ \sqrt{\overline{E^2}} = 0{,}024 \text{ СГСЭ} = 7{,}2 \text{ В/см}.$$

§ 6. Предельный переход от волновой оптики к геометрической

1. Геометрическая оптика является приближенным предельным случаем, в который переходит волновая оптика, когда длина световой волны стремится к нулю. Чтобы показать это, надо было бы исходить из уравнений Максвелла в неоднородных средах. Однако такой путь приводит к громоздким вычислениям. Мы поступим иначе. Среду, в которой распространяется свет, будем считать прозрачной и однородной. Предполагая сначала, что она изотропна, исключим из уравнений (5.1) и (5.2) вектор **H**. С этой целью первое уравнение (5.1) дифференцируем по t, а от обеих частей второго возьмем операцию rot, воспользовавшись при этом векторной формулой

$$\operatorname{rot}\operatorname{rot}\mathbf{E} = \operatorname{grad}\operatorname{div}\mathbf{E} - \Delta\mathbf{E} \ ^{1)}, \tag{6.1}$$

где Δ — оператор Лапласа в прямоугольной системе координат, т.е.

$$\Delta \equiv \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{6.2}$$

---

[1] Эту формулу легко получить, записав левую часть в виде $[\nabla[\nabla\mathbf{E}]]$ и раскрыв по обычному правилу двойное векторное произведение. При этом надо только помнить, что векторы ∇ и $\mathbf{E}$ нельзя переставлять. Таким путем получаем: $\nabla(\nabla\mathbf{E}) - \nabla^2\mathbf{E}$, т.е. правую часть (6.1).

Из полученных таким образом соотношений легко исключить **H**. В результате получится

$$\Delta \mathbf{E} - \frac{1}{v^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \tag{6.3}$$

где v определяется прежним выражением (5.9). Уравнение (6.3) называется *волновым*. Такому же уравнению удовлетворяет и вектор **H**.

Для неоднородных сред уравнение (6.3) усложняется. Однако если интересоваться *только интенсивностью волн*, отвлекаясь от их *поляризации*, то оказывается, что в предельном случае геометрической оптики уравнение (6.3) приводит к правильным результатам [1]. Поэтому даже в случае неоднородных сред предельный переход к геометрической оптике можно выполнить на основе волнового уравнения

$$\Delta E - \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2} = 0, \tag{6.4}$$

в котором E означает длину вектора **E**, а скорость v считается известной функцией координат. Хотя такой путь и не вполне удовлетворителен, но на нем проще уяснить метод, применяемый при обосновании геометрической оптики. Результаты, к которым мы придем, применимы не только к световым, но и *ко всем другим волнам*, например акустическим или волнам де Бройля в квантовой механике.

2. Предполагая волну монохроматической, запишем ее в виде

$$E = a(\mathbf{r})e^{i(\omega t - k_0\Phi)}, \tag{6.5}$$

где $a(\mathbf{r})$ и $\Phi(\mathbf{r})$ — вещественные функции координат. Волновое число в вакууме $k_0 = \omega/c = 2\pi/\lambda$ введено для удобства, как большой размерный параметр. Подставим выражение (6.5) в уравнение (6.4) и отделим вещественную часть от мнимой. В результате получим два уравнения:

$$(\operatorname{grad}\Phi)^2 = n^2 + \frac{\Delta a}{k_0^2 a}, \tag{6.6}$$

$$a\Delta\Phi + 2\operatorname{grad} a \operatorname{grad}\Phi = 0. \tag{6.7}$$

Допустим теперь, что длина волны мала, а амплитуда a меняется в пространстве не очень быстро, так что соблюдается неравенство

$$\left|\frac{\Delta a}{k_0^2 a}\right| \equiv \frac{\lambda^2}{4\pi^2}\left|\frac{\Delta a}{a}\right| \ll n^2. \tag{6.8}$$

---

[1] Предельный переход к геометрической оптике на основе векторных уравнений Максвелла подробно исследован в книге: *Сивухин Д.В.* Лекции по физической оптике, ч. II. Ротапринтное издание. — Новосибирск, 1969.

Для этого достаточно, чтобы выполнялись условия

$$\left|\lambda\frac{\partial a}{\partial x}\right| \ll a, \qquad \left|\lambda\frac{\partial^2 a}{\partial x^2}\right| \ll \left|\frac{\partial a}{\partial x}\right| \tag{6.9}$$

для любого направления оси X. Действительно, тогда

$$\left|\lambda^2\frac{\partial^2 a}{\partial x^2}\right| \ll \left|\lambda\frac{\partial a}{\partial x}\right| \ll a,$$

что совпадает с (6.8), так как $|\Delta a| \sim |\partial^2 a/\partial x^2|$. Пренебрегая в (6.6) последним членом, получим

$$(\operatorname{grad}\Phi)^2 = n^2. \tag{6.10}$$

Уравнения (6.10) и (6.7) и составляют *систему уравнений геометрической оптики.* Из их вывода ясно, что *условием применимости геометрической оптики является малость изменения амплитуды волны и ее первых пространственных производных на протяжении длины волны.* В противном случае могут возникать заметные отступления от геометрической оптики. Это происходит, например, в следующих случаях: 1) на границе геометрической тени; 2) вблизи фокуса, т.е. геометрической точки схождения лучей; 3) при распространении света в среде с резко меняющимся показателем преломления (например, в мутной среде); 4) при распространении света в сильно поглощающих средах (например, металлах).

3. Величину Φ Клаузиус (1822–1888) назвал *эйконалом*, а уравнение (6.10) — *уравнением эйконала.* Его можно записать в векторной форме:

$$\operatorname{grad}\Phi = n\mathbf{s}, \tag{6.11}$$

где $\mathbf{s}$ — единичный вектор нормали к фронту волны

$$\omega t - k_0\Phi = \text{const}, \tag{6.12}$$

проведенный в сторону ее распространения.

Уравнение эйконала определяет скорость распространения волнового фронта в направлении нормали $\mathbf{s}$. Действительно, на основании определения градиента можно записать (6.11) в виде $\partial\Phi/\partial s = n$. С другой стороны, дифференцирование уравнения распространения волнового фронта (6.12) дает $\omega\,dt = k_0\,d\Phi$, или

$$\omega\,dt = \frac{\omega}{c}\frac{\partial\Phi}{\partial s}\,ds = \frac{\omega}{c}\,n\,ds = \frac{\omega}{v}\,ds.$$

Отсюда для нормальной скорости волнового фронта находим

$$\frac{ds}{dt} = v, \tag{6.13}$$

т.е. эта скорость такая же, как у плоской волны. Этого и следовало ожидать, так как малый участок волнового фронта в малых объемах пространства должен вести себя как плоский.

Полученный результат позволяет построить волновой фронт F_2 в момент времени $t + dt$, если известно его положение F_1 в момент t. Для этого из каждой точки исходного волнового фронта F_1 (рис. 19) следует отложить в направлении нормали отрезок длиной $v\,dt$. Соединив концы всех таких отрезков, мы и получим волновой фронт F_2 в момент $t + dt$. Вместо этого можно из каждой точки волнового фронта F_1, как из центра, описать сферы радиусом $v\,dt$. Огибающая таких сфер и будет волновым фронтом F_2. Оба построения совершенно эквивалентны. Тем самым построение Гюйгенса (см. § 3, п. 4) распространяется и на волны в неоднородных средах.

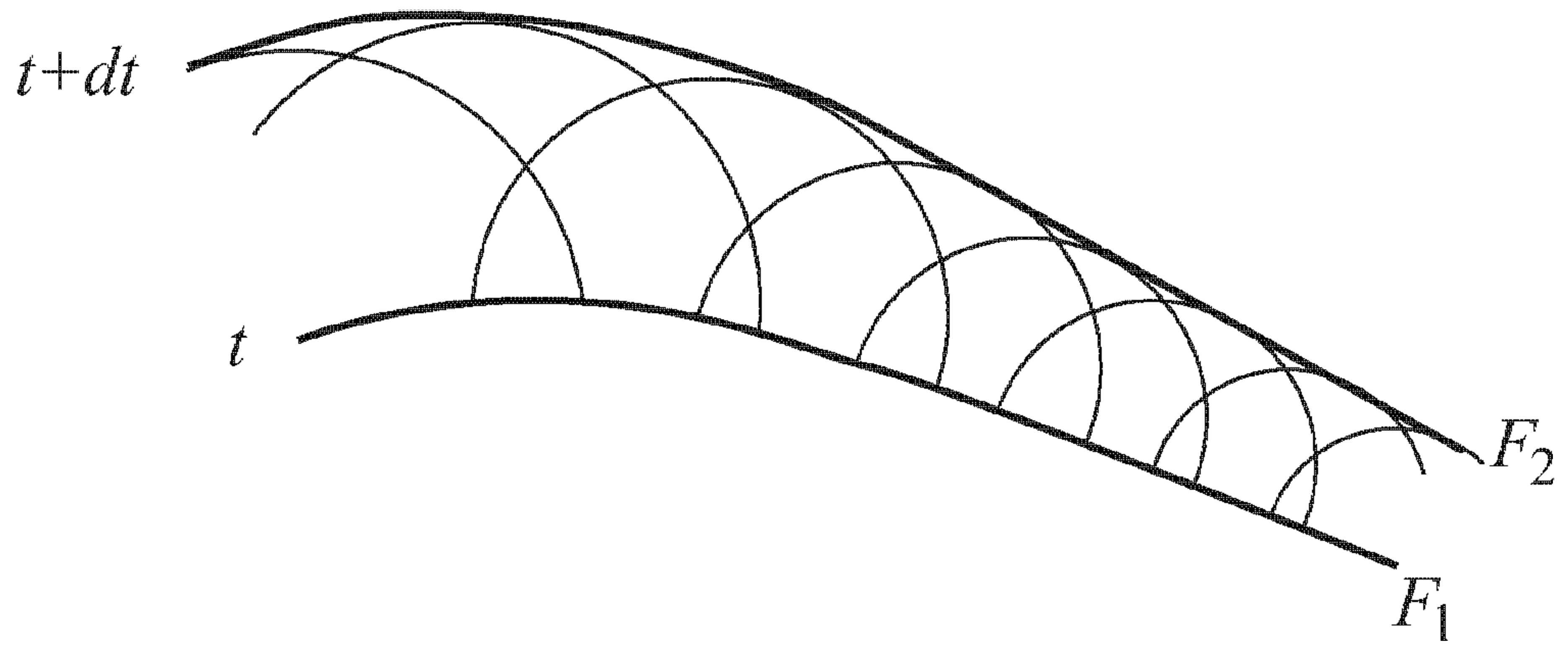


Рис. 19

4. Второе уравнение геометрической оптики (6.7) теперь можно записать в виде

$$a\Delta\Phi + 2n\frac{\partial a}{\partial s} = 0. \tag{6.14}$$

Если определить *луч как ортогональную траекторию к семейству волновых фронтов, или семейству равных фаз* (6.12), то взятие производной по s можно понимать в смысле дифференцирования по длине луча s. Интегрируя уравнение (6.14) вдоль луча, придем к соотношению

$$a = a_0 \exp\left(-\int_0^s \frac{\Delta\Phi}{2n}\,ds\right), \tag{6.15}$$

где a_0 — амплитуда в «начальной точке» луча, от которой отсчитывается длина s. Формула (6.15) показывает, что для определения волнового поля во всех точках луча достаточно знать его значение в какой-либо одной точке того же луча. Но уравнения геометрической оптики ничего не могут сказать относительно изменения амплитуды поля при переходе от одного луча к соседнему. Они допускают любые изменения амплитуды от луча к лучу. Необходима только достаточная медленность такого изменения, чтобы волна, формально удовлетворяющая уравнениям геометрической оптики, могла быть реализована в действительности. Таким образом, *в приближении геометрической оптики* *световое поле на всяком луче совершенно не зависит от полей других лучей.*

Отсюда следует основное представление геометрической оптики о *распространении световой энергии вдоль лучей*, точнее — вдоль «лучевых трубок», образованных лучами. Отсюда также следует, что при нахождении волнового фронта построением Гюйгенса новый волновой фронт не выходит за пределы огибающей вторичных сферических волн Гюйгенса. Тем самым дано полное обоснование гипотезы Гюйгенса об огибающей и показано, что эта гипотеза справедлива только в приближении геометрической оптики.

К представлению о распространении света вдоль лучей можно прийти и другим путем. Умножая уравнение (6.7) на a и замечая, что $\Delta\Phi = \operatorname{div}\operatorname{grad}\Phi = \operatorname{div}(n\mathbf{s})$, перепишем это уравнение в виде

$$a^2\operatorname{div}(n\mathbf{s}) + 2a\operatorname{grad}a(n\mathbf{s}) = 0,$$

или

$$\operatorname{div}(na^2\mathbf{s}) = 0. \tag{6.16}$$

Полученное соотношение по форме совпадает с уравнением непрерывности $\operatorname{div}\mathbf{j} = 0$ для стационарного течения несжимаемой жидкости. Роль линий тока играют световые лучи, а плотности потока жидкости — вектор $\mathbf{j} = na^2\mathbf{s}$, пропорциональный плотности потока световой энергии. Свет как бы течет вдоль узких «световых трубок», т.е. трубок, боковые стенки которых образованы лучами. Через эти боковые стенки свет не проникает. Если σ — поперечное сечение трубки, то вдоль нее величина $na^2\sigma$ сохраняется неизменной, как это видно из уравнения (6.16).

Можно выделить отдельную световую трубку, или *физический световой луч*, поставив на пути распространяющейся волны (6.5) узкую диафрагму. Только диафрагма не должна быть особенно узкой, а световая трубка слишком длинной. Дело в том, что на краях диафрагмы и вблизи боковых границ трубки амплитуда поля меняется резко, т.е. условия применимости геометрической оптики не выполняются. Возникает дифракция света, приводящая к уширению светового пучка. Однако если диафрагма не слишком мала, а световая трубка не слишком длинна, эти эффекты малосущественны. Но они всегда скажутся на больших расстояниях от диафрагмы. В теории дифракции будет показано, что необходимым условием, при выполнении которого можно говорить о физическом световом луче, является неравенство

$$l \ll D^2/\lambda, \tag{6.17}$$

где D — минимальный линейный размер диафрагмы, а l — расстояние от диафрагмы, измеренное вдоль луча.

Оптической длиной линии в однородной среде называется произведение геометрической длины этой линии l на показатель преломления n. Если среда неоднородна, то оптическая длина определяется интегралом $\int n\,dl$, взятым вдоль рассматриваемой линии. Если ABC — геометрическая длина линии, то оптическая длина ее обозначается через (ABC), т.е. заключением в круглые скобки геометрической длины. Из построения Гюйгенса, изложенного в п. 3 этого параграфа, следует, что оптические длины всех лучей между двумя положениями волнового фронта равны между собой.

5. Нетрудно получить из волновых представлений и выражение для кривизны луча в однородной среде. Для этого через бесконечно малый отрезок луча AC проведем соприкасающуюся плоскость (рис. 20). Она пересечет волновые фронты, проходящие через концы этого отрезка, вдоль кривых AB и CD. Пусть BD — бесконечно близкий луч, лежащий в той же плоскости. Так как лучи перпендикулярны к волновым фронтам, то все углы бесконечно малого криволинейного четырехугольника $ABDC$ прямые. А так как оптические длины лучей между любыми двумя положениями волнового фронта одинаковы, то $nl = (n+dn)(l+dl)$, где l и $l+dl$ — длины отрезков AC и BD, а n и $n+dn$ — соответствующие им показатели преломления. С точностью до бесконечно малых высшего порядка отсюда получаем $l\,dn + n\,dl = 0$. По определению радиуса кривизны $l = R\varphi$, где φ — угол между касательными к лучу AC в точках A и C, равный углу между касательными к отрезкам AB и CD в тех же точках. Очевидно, $dl = -a\varphi$, где a — длина отрезка AB. Приращение показателя преломления dn происходит вдоль главной нормали $\mathbf{N}$, так что $dn = (\partial n/\partial N)a$. Учитывая все это, получаем: $R\varphi(\partial n/\partial N)a - na\varphi = 0$, откуда

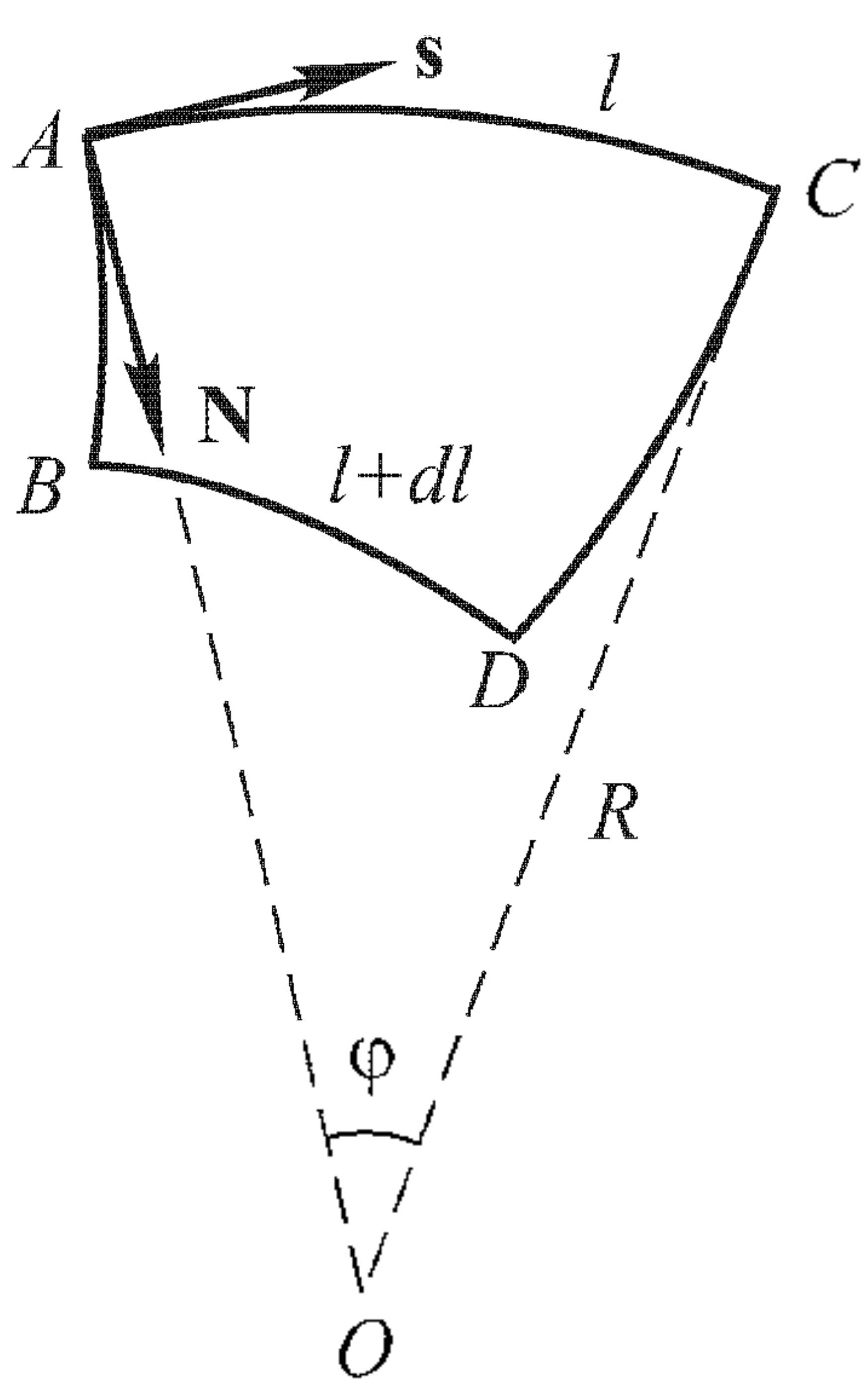


Рис. 20

$$\frac{1}{R} = \frac{1}{n}\frac{\partial n}{\partial N}, \tag{6.18}$$

что совпадает с формулой (4.1).

§ 7. Принцип Ферма

1. Пьер Ферма (1601–1675) выдвинул принцип, согласно которому *свет при распространении из одной точки в другую выбирает путь, которому соответствует наименьшее время распространения.* Ферма руководствовался телеологическими соображениями, согласно которым природа действует целенаправленно: она не может быть расточительной и должна достигать своих целей с наименьшей затратой средств. Подобные соображения, конечно, чужды науке и не могут служить обоснованием принципа Ферма. Но сам принцип (после введения некоторых уточнений) верен и может оказаться полезным при решении отдельных вопросов геометрической оптики. Это было продемонстрировано уже самим Ферма, который с помощью своего принципа вывел закон преломления Снеллиуса и получил такое же выражение для показателя преломления, что и в волновой теории света. В частности, он пришел к заключению, что скорость света в более преломляющей среде меньше, чем в менее преломляющей.

2. Для доказательства принципа Ферма допустим сначала, что показатель преломления среды меняется в пространстве непрерывно и достаточно медленно, так что условия применимости геометрической оптики выполнены. Пусть в среде распространяется волна вида (6.5), например порожденная точечным источником. Ей соответствует система лучей, представленная на рис. 21. Если эйконал Φ — однозначная функция координат, то из уравнения (6.11) следует, что циркуляция вектора $n\mathbf{s}$ по любому замкнутому контуру равна нулю, т.е.

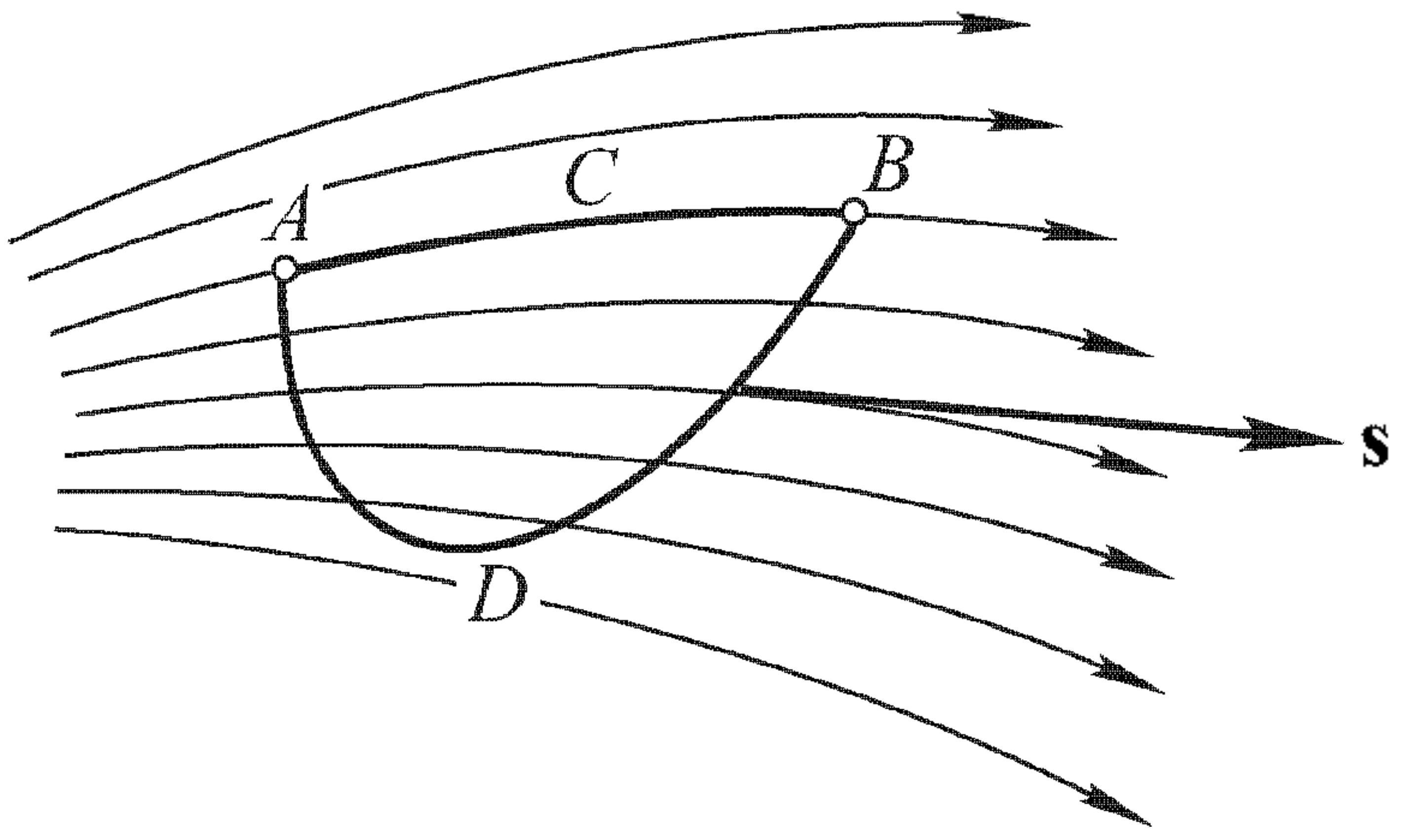


Рис. 21

$$\oint n(\mathbf{s}\,d\mathbf{l}) = 0, \qquad (7.1)$$

где $d\mathbf{l}$ — вектор элементарного смещения вдоль этого контура. Возьмем две произвольные точки A и B, лежащие на одном из лучей. Соединим их произвольной линией ADB. В силу (7.1)

$$\int\limits_{ACB} n(\mathbf{s}\,d\mathbf{l}) = \int\limits_{ADB} n(\mathbf{s}\,d\mathbf{l}).$$

На луче ACB векторы $\mathbf{s}$ и $d\mathbf{l}$ направлены одинаково, следовательно, $(\mathbf{s}\,d\mathbf{l}) = dl$. На линии же ADB $(\mathbf{s}\,d\mathbf{l}) = dl\cos(\mathbf{s}, d\mathbf{l}) \leqslant dl$. Поэтому

$$\int\limits_{ACB} n\,dl \leqslant \int\limits_{ADB} n\,dl. \qquad (7.2)$$

Знак равенства относится только к случаю, когда кривая ADB сама является лучом. Таким образом, *если показатель преломления меняется в пространстве непрерывно, то оптическая длина луча между любыми двумя точками меньше оптической длины всякой другой линии, соединяющей те же точки.* Но это есть другая формулировка принципа Ферма, так как оптическая длина луча пропорциональна времени распространения света вдоль него.

Приведенная формулировка принципа Ферма нуждается в уточнении. В некоторых случаях она может оказаться неверной. Рассмотрим, например, среду со сферически симметричным распределением показателя преломления вокруг центра O (рис. 22). Примером такой среды может служить планетная атмосфера. Предположим, что показатель преломления меняется в пространстве так, что световой луч, выйдя из какой-либо точки перпендикулярно к радиусу, описывает окружность с центром в точке O. Пусть свет попадает из точки A в точку B по большой дуге ACB этой окружности. Но он может пройти из A в B и по дуге ADB той же окружности, затрачивая на распространение меньшее время. Меньшее время потребовалось бы и в том случае, если бы свет избрал какой-либо другой путь, бесконечно близкий к дуге ADB. Все это противоречит принципу Ферма в приведенной выше формулировке.

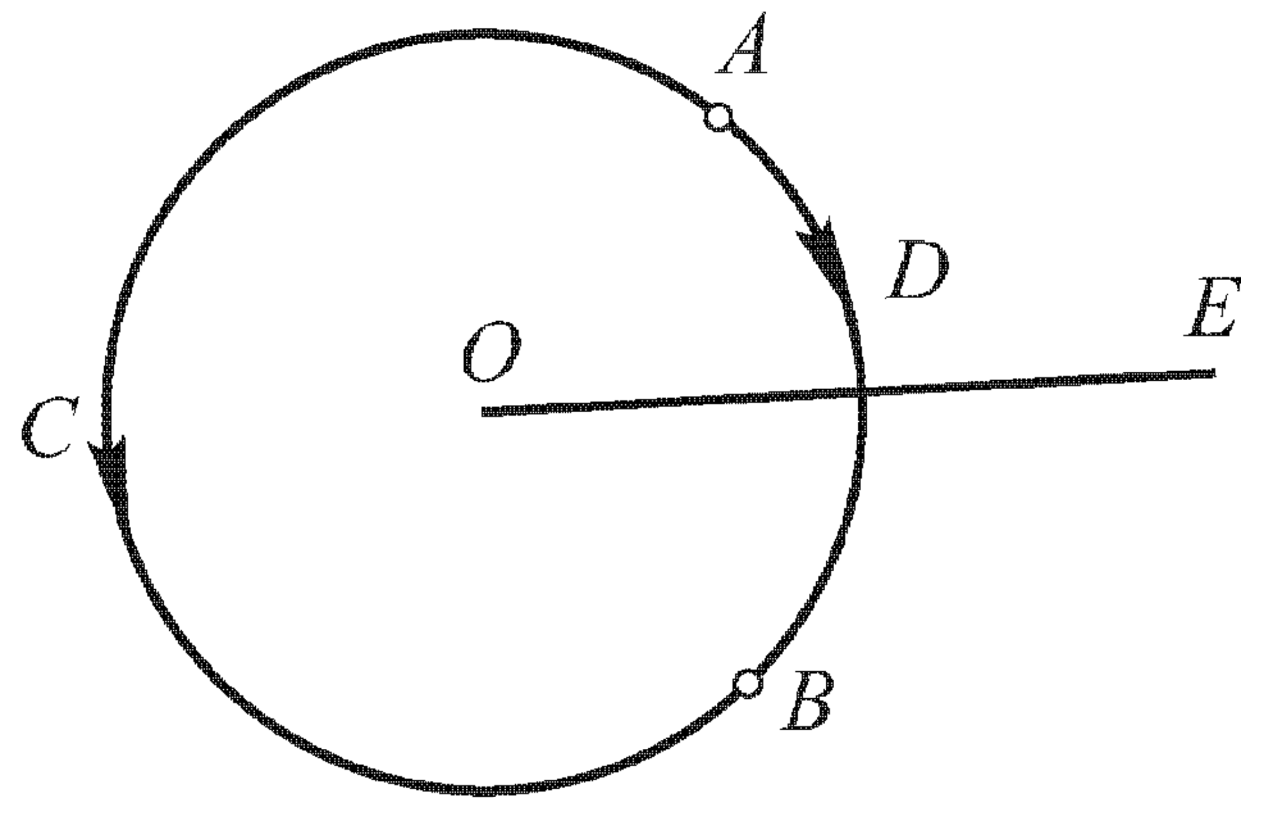


Рис. 22

Причина противоречия состоит в том, что в приведенном примере эйконал Φ не есть однозначная функция координат, как это предполагалось при выводе. Действительно, если луч описывает окружность вокруг центра O, то он вернется в исходную точку с новым значением эйконала: эйконал Φ получит приращение nl, где l — длина описанной окружности. Если окружность описывается m раз, то приращение эйконала будет $2mnl$. Это и значит, что функция Φ неоднозначна. Для справедливости принципа Ферма необходимо наложить на выбор воображаемых путей распространения света такие ограничения, чтобы эйконал Φ вел себя как *однозначная функция координат*. В приведенном примере этого можно достигнуть, поставив перегородку вдоль меридиональной полуплоскости ODE и ограничиваясь только такими путями, которые не пересекают эту перегородку.

Подобным приемом можно воспользоваться и во всех остальных случаях, в которых эйконал Φ оказывается неоднозначным. Впрочем, в применениях принципа Ферма достаточно ограничиться только такими путями, которые проходят бесконечно близко от действительного пути света. В этом случае надобность во введении перегородок отпадает.

3. При наличии поверхностей раздела сред, на которых лучи могут испытывать отражение или преломление, в формулировку и доказательство принципа Ферма надо ввести дополнения. Пусть луч, выйдя из точки A (рис. 23), после отражений или преломлений в точках C, D, E, ... попадает в точку B. Назовем *виртуальным путем света* любую линию $AC'D'E'B$ между крайними точками A и B, которая получается из $ACDEB$ в результате бесконечно малого бокового смещения ее и отличается от нее бесконечно мало по направлению. Принцип Ферма утверждает, что *оптическая длина действительного светового пути* (*или пропорциональное ей время распространения*) *стационарна.* Это значит, что разность оптических длин действительного и виртуального путей света есть величина *более высокого порядка малости*, чем боковое смещение виртуального пути относительно действительного. Только эта стационарность, а не минимальность оптической длины луча и существенна в приложениях.

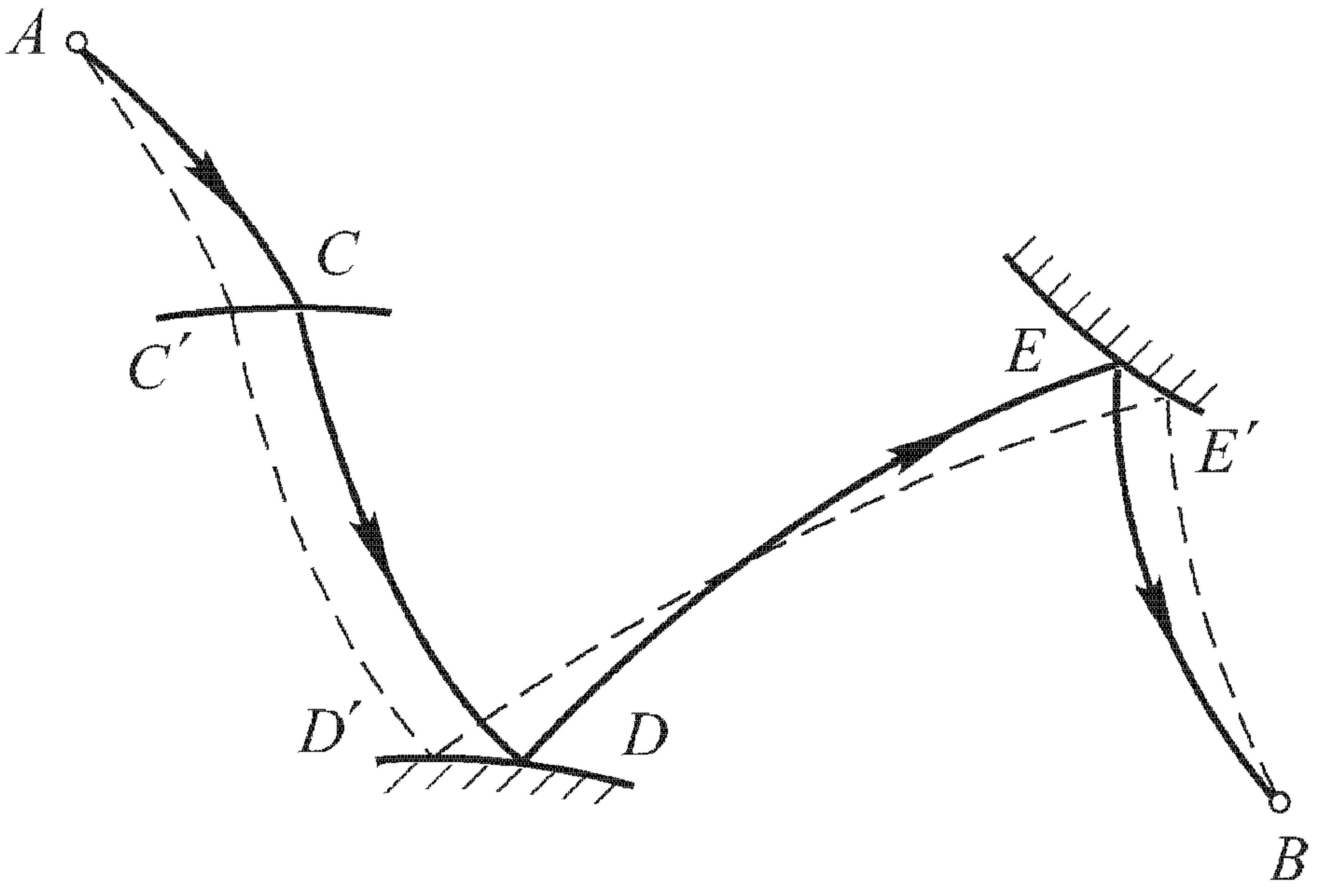


Рис. 23

При доказательстве достаточно ограничиться преломлением на одной границе. Случай отражения исследуется так же. Пусть MN — граница раздела сред *1* и *2*, а ACB — действительный луч, соединяющий точку A с точкой B (рис. 24). Вообразим два бесконечно узких пучка лучей: один в первой среде, исходящий из точки A, другой во второй среде, сходящийся в точке B. За положительные направления лучей примем направления от A к B. Выберем в этих пучках два луча AC' и $C'B$, пересекающихся на границе раздела в точке C'. Кривую $AC'B$ можно рассматривать как виртуальный путь света, так как луч $C'B$ в общем случае, отнюдь не возникает в результате преломления луча AC'. Обозначим через Φ_1 и Φ_2 эйконалы рассматриваемых пучков лучей, отсчитываемые от точек A и B соответственно. Тогда

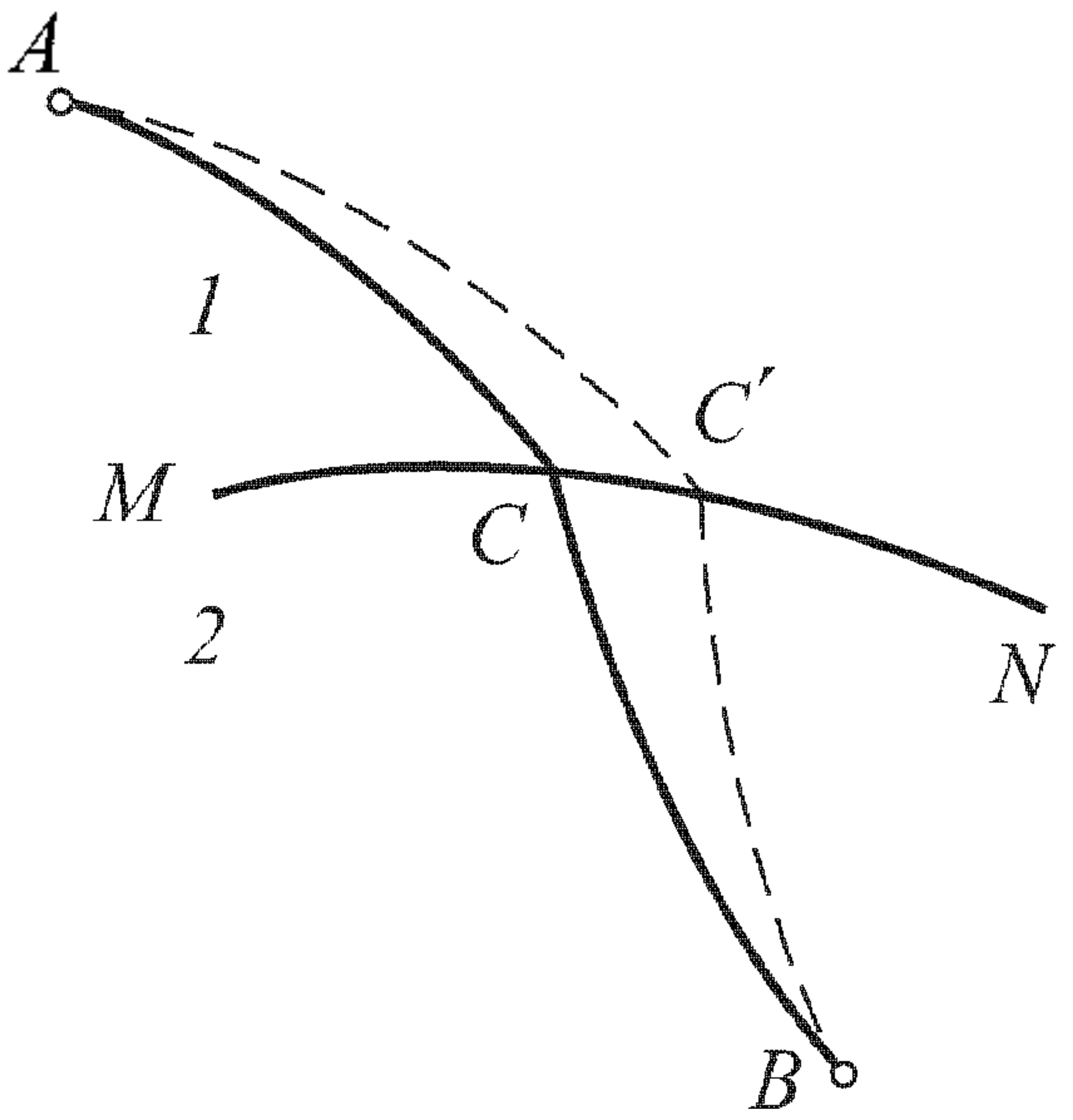


Рис. 24

$$\int\limits_{ACB} n\,ds = \int\limits_{AC} n\,ds + \int\limits_{CB} n\,ds = \int\limits_{AC} n\,ds - \int\limits_{BC} n\,ds = \Phi_1(C) - \Phi_2(C).$$

Вариация интеграла $\int n\,ds$ при смещении точки C в произвольную бесконечно близкую точку C' границы раздела будет

$$\delta \int n\,ds = \delta\Phi_1 - \delta\Phi_2.$$

Если $\delta\mathbf{r} \equiv \overrightarrow{CC'}$ — вектор смещения, то $\delta\Phi_1 = (\operatorname{grad}\Phi_1\delta\mathbf{r}) = n_1(\mathbf{s}_1\delta\mathbf{r})$ и аналогично $\delta\Phi_2 = n_2(\mathbf{s}_2\delta\mathbf{r})$, так что

$$\delta \int n\,ds = (n_1\mathbf{s}_1 - n_2\mathbf{s}_2)\delta\mathbf{r}.$$

В силу закона преломления Снеллиуса вектор $(n_1\mathbf{s}_1 - n_2\mathbf{s}_2)$ перпендикулярен к границе раздела сред в точке падения, а потому и к бесконечно малому смещению вдоль границы $\delta\mathbf{r}$. Таким образом, в первом порядке по δr вариация оптической длины луча ACB обращается в нуль. При доказательстве предполагалось, что виртуальный путь состоит из отрезков лучей AC' и $C'B$. Однако результат не изменится, если эти отрезки заменить произвольными бесконечно близкими к ним линиями, соединяющими те же точки A и C', C' и B. В самом деле, поскольку AC' и $C'B$ — действительные лучи в первой и второй средах, их оптические длины по доказанному выше минимальны. По этой причине замена действительных лучей AC' и $C'B$ бесконечно близкими к ним линиями, соединяющими те же крайние точки, не меняет в первом порядке оптические длины соответствующих путей. Следовательно, вариация оптической длины луча ACB останется равной нулю, каков бы ни был виртуальный путь света. А к этому в рассматриваемом случае и сводится содержание принципа Ферма.

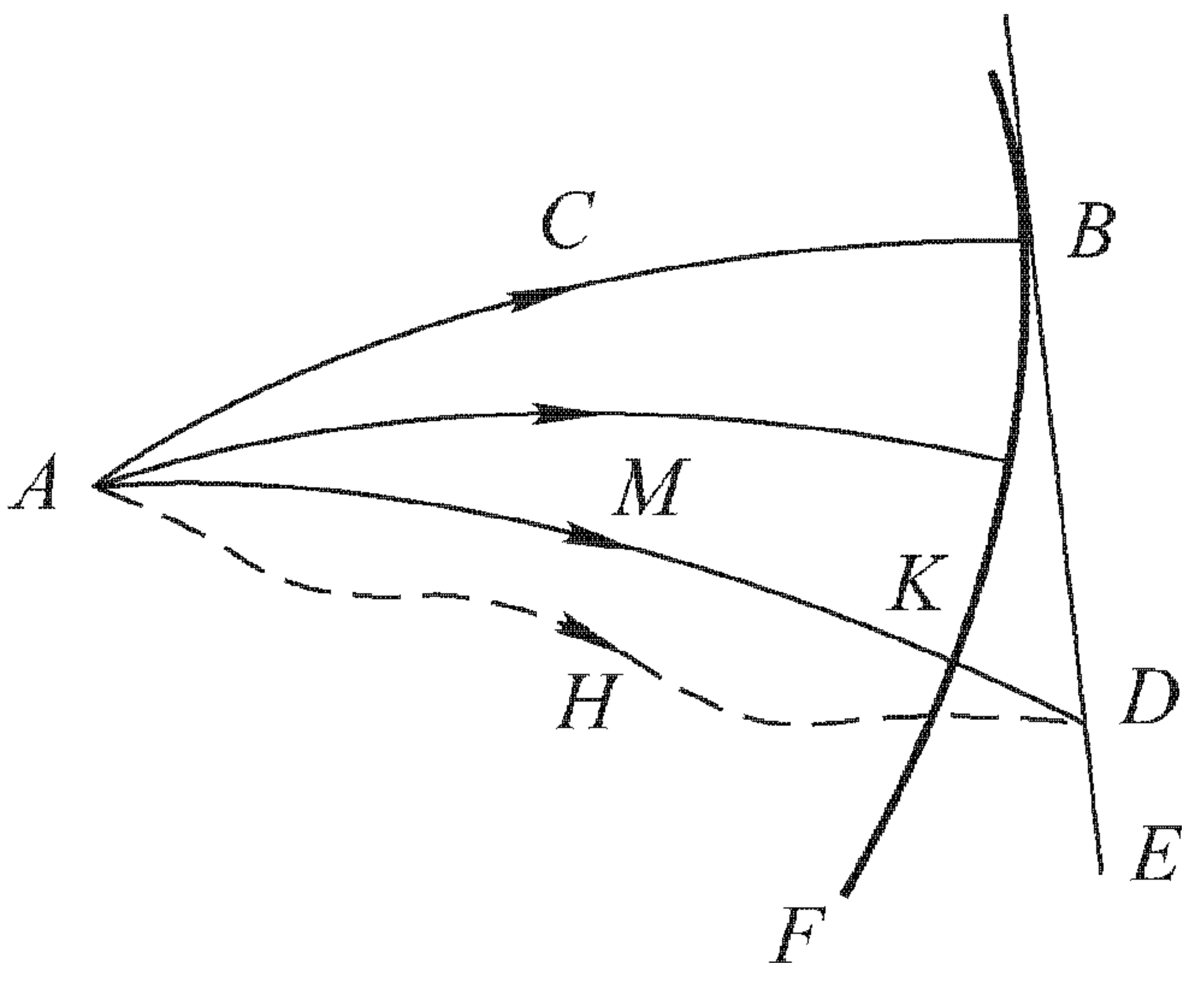


Рис. 25

4. В применениях иногда удобна следующая теорема, являющаяся непосредственным следствием принципа Ферма. *Пусть A и B — произвольные точки луча ACB* (рис. 25). *Проведем через точку B произвольную гладкую поверхность BE, ортогональную к лучу ACB в точке B. Пусть BD — бесконечно малое смещение вдоль этой поверхности. Соединим начальную точку луча A с точкой D произвольной линией AHD, бесконечно мало отличающейся по направлению от луча ACB. Тогда вариация оптической длины при переходе от истинного пути света ACB к виртуальному AHD будет равна нулю.*

Для доказательства возьмем пучок лучей, исходящих из точки A. Все эти лучи ортогональны к волновому фронту BF, а их оптические длины от точки A до волнового фронта одинаковы. В частности, $(ACB) = (AMK)$. Но по принципу Ферма с точностью до бесконечно малых высшего порядка $(AMK) = (AHK)$. Далее, поскольку поверхности BDE и BKF касаются друг друга в точке B, длина луча KD будет бесконечно малой высшего порядка по сравнению с BD. Поэтому оптическая длина AHD будет отличаться от оптической длины ACB также на величину высшего порядка малости по сравнению с боковым смещением BD. Это и требовалось доказать.

5. Если свет распространяется в однородных средах, граничащих между собой, то в каждой среде путь света будет прямолинеен. В этом случае задача сводится только к нахождению точек на поверхностях раздела сред, в которых происходит отражение и преломление светового луча. Поэтому нет необходимости вводить криволинейные виртуальные пути света. Достаточно ограничиться ломаными виртуальными путями, состоящими из отрезков прямых линий, причем изломы таких путей должны происходить на границах раздела рассматриваемых сред. Даже при таких ограничениях оптическая длина действительного светового пути может быть не только минимальной, но и максимальной или стационарной.

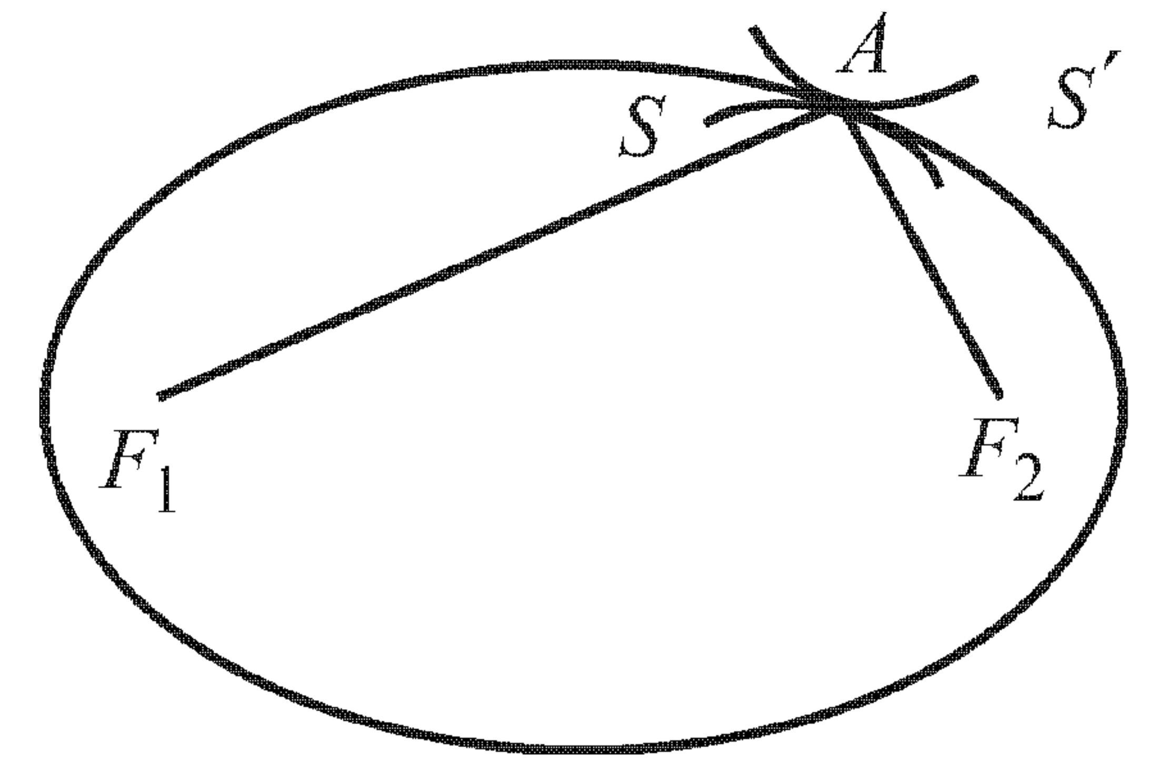


Рис. 26

Чтобы показать это в случае отражения света, возьмем эллипсоидальное зеркало, получающееся от вращения эллипса вокруг его большой оси F_1F_2 (рис. 26).

Пусть F_1 и F_2 — фокусы эллипсоида. Если A — точка на его поверхности, то $F_1A + F_2A = 2a$, где $2a$ — длина большой оси эллипсоида. Поверхность зеркала делит все пространство на две части: *внутреннюю*, сумма расстояний каждой точки которой от фокусов F_1 и F_2 меньше $2a$, и *внешнюю*, для которой эта сумма больше $2a$. Пусть световой луч выходит из фокуса F_1. Тогда после отражения от эллипсоидального зеркала в точке A он пройдет через второй фокус F_2, так как по известному свойству эллипса прямые F_1A и F_2A образуют одинаковые углы с нормалью к поверхности зеркала. При смещении вдоль поверхности зеркала сумма $F_1A + F_2A$, а с ней и время распространения света из F_1 в F_2 не изменяются. Вариация времени распространения при таком смещении равна нулю. Однако это время ни минимально, ни максимально — оно *постоянно*. Именно по этой причине любой луч, вышедший из F_1, обязательно пройдет через F_2, в какой бы точке зеркала он ни отразился. Убедиться в этом можно с помощью таких же рассуждений, какие были приведены в п. 3.

Вообразим теперь зеркало S, касающееся эллипсоида в точке A, обращенное вогнутостью в ту же сторону, что и эллипсоид, но имеющее бо́льшую кривизну. Световой луч F_1A после отражения от этого зеркала снова попадает в точку F_2. Однако при смещении точки A по поверхности зеркала S длина ломаной F_1AF_2 уменьшается. Следовательно, время распространения света из F_1 в F_2 вдоль действительного пути *максимально.* Наоборот, если взять зеркало S', имеющее в точке касания меньшую кривизну, чем эллипсоид, или обращенное вогнутостью в противоположную сторону, то время распространения света вдоль действительного пути будет *минимально.* В частности, оно минимально при отражении от *плоского зеркала.* Допустим, наконец, что зеркало SAS' имеет в A *точку перегиба.* Тогда при смещении точки падения луча по поверхности этого зеркала время распространения либо увеличится, либо уменьшится, либо останется неизменным, в зависимости от направления смещения.

6. Чтобы разобрать случай преломления, введем понятие *анаберрационной поверхности.* Пусть точка P находится в однородной среде с показателем преломления n, а точка P' — в однородной среде с показателем преломления n' (рис. 27). Поверхность AA', вдоль которой среды граничат друг с другом, называется *анаберрационной*, если любая точка A этой поверхности удовлетворяет условию

$$n\cdot PA+n'\cdot AP' = C = \text{const}.$$

Для случая преломления анаберрационная поверхность имеет форму так называемого *картезианского овала* (см. задачу 2 к § 9). Он обращен вогнутостью в сторону более преломляющей среды ($n' > n$). Анаберрационная поверхность делит пространство на две части, обладающие следующим свойством. Если точка M расположена в менее преломляющей среде, то сумма $n\cdot PM + n'\cdot MP'$ больше C; если же она лежит в более преломляющей среде, то эта сумма меньше C.

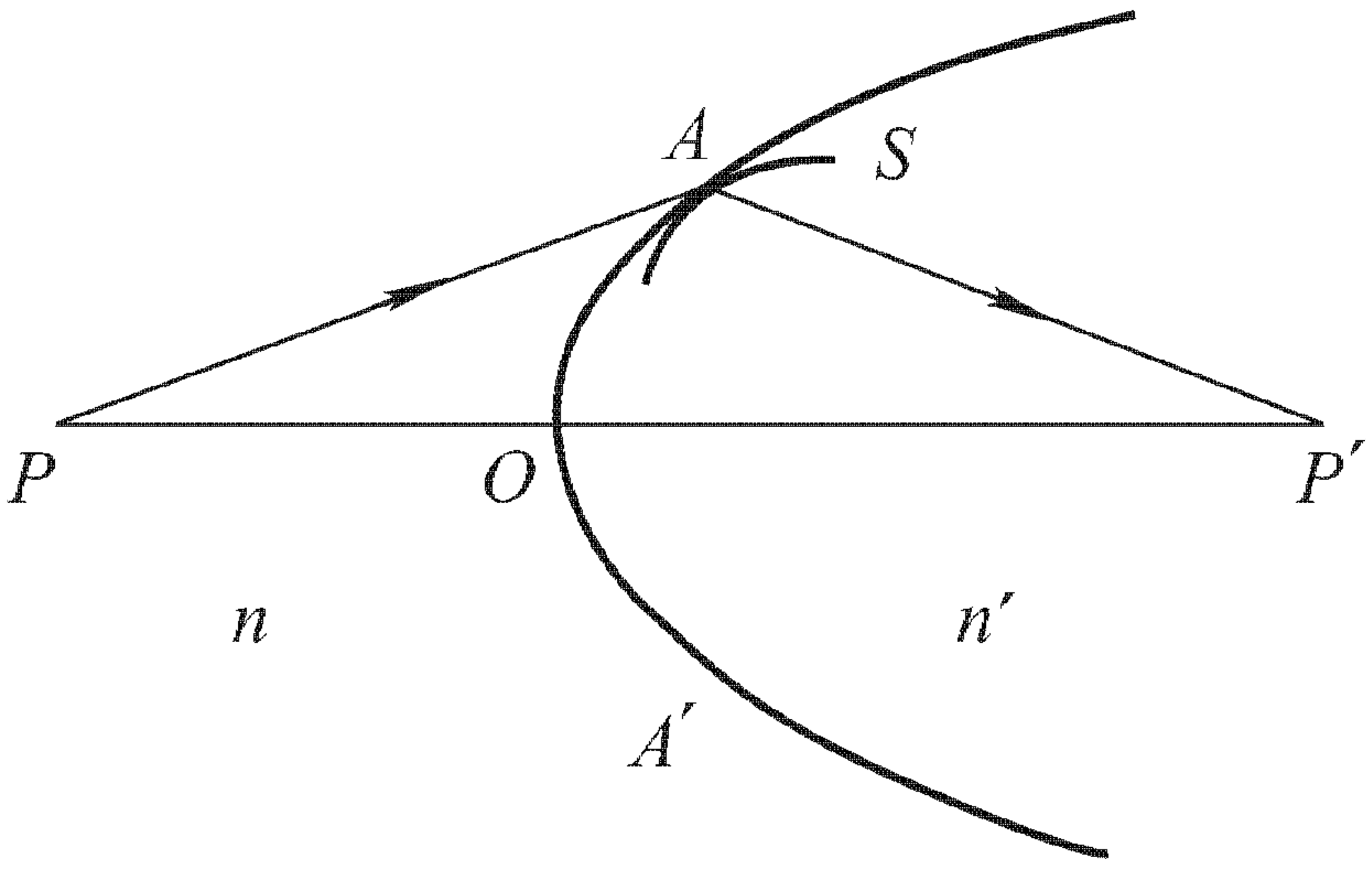


Рис. 27

Докажем следующую теорему. *Луч света, вышедший из точки P, после преломления на анаберрационной поверхности обязательно пройдет через точку P'.* Действительно, пусть PA — падающий луч, а $\mathbf{s}$ — единичный вектор, направленный вдоль него. Соединим точку A с точкой P' и обозначим через $\mathbf{s}'$ единичный вектор, направленный вдоль прямой AP'. По определению анаберрационной поверхности вариация оптической длины ломаной PAP' при смещении точки A по анаберрационной поверхности будет равна нулю. Поэтому, применяя такие же рассуждения, какие были проведены в п. 2, найдем, что вектор $n\mathbf{s} - n'\mathbf{s}'$ перпендикулярен к анаберрационной поверхности в точке A. Отсюда следует, что AP' дает направление преломленного луча.

Доказанной теореме можно дать также следующую формулировку. *Если AA' — анаберрационная поверхность относительно пары точек P и P', то каждая из этих точек будет оптическим изображением другой при преломлении лучей на этой анаберрационной поверхности.* При этом на угловую ширину пучка лучей не накладывается никаких ограничений.

Вернемся к исследованию характера экстремума оптической длины луча при преломлении. Наши рассуждения ничем не будут отличаться от рассуждений, проведенных выше для эллипсоидального зеркала. Допустим, например, что среды граничат друг с другом вдоль поверхности S (рис. 27), касающейся анаберрационной поверхности в точке A. Тогда падающий луч после преломления в точке A опять пройдет через точку P'. Пусть поверхность S обращена вогнутостью в ту же сторону, что и анаберрационная поверхность, и имеет в точке касания бóльшую кривизну. Тогда при смещении точки падения вдоль S она окажется в менее преломляющей среде. Следовательно, смещенный путь будет иметь *меньшую оптическую длину*, чем действительный: время распространения света вдоль действительного пути *максимально.* Напротив, когда кривизна поверхности S в точке касания A меньше кривизны анаберрационной поверхности, а также тогда, когда поверхность S обращена вогнутостью в противоположную сторону, то время распространения вдоль действительного пути *минимально.* В частности, оно минимально при преломлении на *плоской поверхности.*

ЗАДАЧА

Система лучей называется *ортотомной*, если все лучи этой системы ортогональны к одной и той же поверхности. Пользуясь принципом Ферма, доказать *теорему Малюса: ортотомная система лучей остается ортотомной после произвольного числа отражений и преломлений.*

Решение. Пусть все лучи перпендикулярны к поверхности F (рис. 28). Проведем через каждую точку этой поверхности луч и отложим на нем отрезок постоянной (но произвольной) оптической длины L. Геометрическим местом концов таких отрезков будет какая-то поверхность F'. Докажем, что все лучи рассматриваемой системы перпендикулярны к поверхности F', каково бы ни было значение величины L.

Малые отрезки одного из лучей AC и $C'A'$ могут считаться прямолинейными. Возьмем соседний бесконечно близкий луч и притом такой, что длины AB и $A'B'$ бесконечно малы по сравнению с AC и $C'A'$. Соединим B с

C и C' с B прямолинейными отрезками. По принципу Ферма с точностью до бесконечно малых второго или высшего порядков $(BEB') = (BCC'B')$, а по построению $(BEB') = (ACC'A')$. Таким образом, $(BCC'B') = (ACC'A')$. Вычитая отсюда общую часть (CC'), получим: $(AC) + (C'A') = (BC) + (C'B')$. Так как по условию отрезок AC перпендикулярен к AB, то с точностью до второго порядка малости $AC = BC$, а следовательно, $(AC) = (BC)$. Поэтому с той же точностью $(C'A') = (C'B')$, или $C'A' = C'B'$, откуда следует, что $C'A' \perp A'B'$.

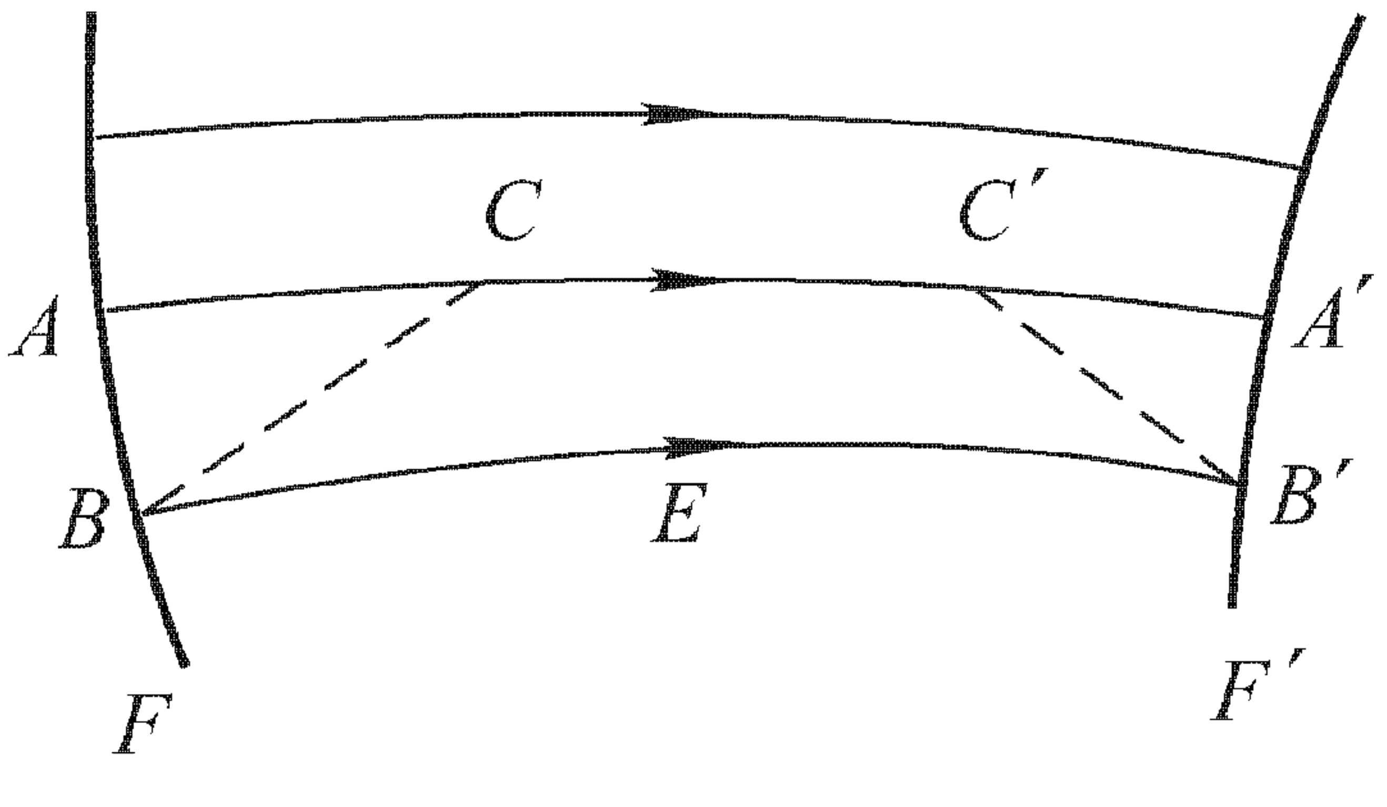


Рис. 28

С точки зрения волновой теории теорема Малюса почти самоочевидна. Действительно, для ортотомной системы лучей поверхность F есть одна из *поверхностей равной фазы* (волновой фронт). Распространяясь по законам геометрической оптики, она продолжает оставаться поверхностью равной фазы, а совокупность лучей — *ортогональной системой*. Конечно, ортогональность может и не соблюдаться. Например, волны вида (6.5) при соблюдении принципа суперпозиции распространяются независимо друг от друга. Каждой из таких волн соответствует ортотомная система лучей. Однако совокупность лучей, соответствующих *всем волнам*, ортотомную систему, вообще говоря, не образует.

§ 8. Групповая скорость

1. До сих пор при рассмотрении скорости распространения волн мы предполагали, что соблюдается принцип суперпозиции и отсутствует дисперсия. При несоблюдении принципа суперпозиции и наличии дисперсии вопрос о скорости распространения волн становится очень сложным. *Ниже предполагается, что принцип суперпозиции соблюдается, но имеется дисперсия.* Сначала рассмотрим плоские волны, распространяющиеся в одном направлении, принимаемом за направление оси X.

Бегущую плоскую монохроматическую волну запишем в виде

$$E = E_0 \cos(\omega t - kx + \delta), \tag{8.1}$$

где E_0 и δ — постоянные. При рассмотрении таких волн дисперсия не играет роли, поскольку частота ω имеет единственное значение. Для выяснения смысла скорости распространения рассмотрим уравнение

$$\omega t - kx + \delta = \text{const}. \tag{8.2}$$

Это есть уравнение плоскости, перпендикулярной к оси X, на которой постоянна фаза волны. Дифференцируя его, получим:

$\omega\, dt - k\, dx = 0$, откуда

$$\frac{dx}{dt} = \frac{\omega}{k}. \tag{8.3}$$

Таким образом, ω/k есть *скорость распространения поверхности постоянной фазы*, ранее обозначавшаяся через v. Она называется *фазовой скоростью волны*. С такой скоростью распространяется синусоидальная волна типа (8.1) без изменения своей формы.

Если бы среда не обладала дисперсией, то говорить о какой-либо другой скорости распространения волны не было бы необходимости. Действительно, произвольное плоское возмущение, распространяющееся в направлении оси X, согласно теореме Фурье можно представить в виде суперпозиции монохроматических волн вида (8.1). При отсутствии дисперсии все эти волны имели бы одну и ту же фазовую скорость, так что форма возмущения все время оставалась бы *одной и той же*. Возмущение в целом бежало бы вперед без изменения вида со скоростью, равной той же фазовой скорости. Не то будет при наличии дисперсии. Тогда монохроматические волны разных частот будут распространяться вперед с разными скоростями, вследствие чего форма всего возмущения будет *непрерывно деформироваться* (исключение составляют только монохроматические волны, распространяющиеся по-прежнему без изменения формы). В этих условиях понятие скорости распространения утрачивает тот ясный смысл, какой оно имело при отсутствии дисперсии. При определенных условиях, однако, можно сохранить представление о скорости распространения немонохроматических волн и в средах, обладающих дисперсией. Важнейшей после фазовой скорости является так называемая *групповая скорость*.

2. Рассмотрим сначала в действительности не существующую непоглощающую среду, в которой фазовая скорость v выражается линейной функцией длины волны λ:

$$v = a + b\lambda, \tag{8.4}$$

а следовательно, частота ω — линейной функцией волнового числа k:

$$\omega = ak + 2\pi b. \tag{8.5}$$

Произвольное плоское возмущение, распространяющееся в среде, разложим на монохроматические волны. Их число, вообще говоря, будет бесконечно велико. Однако можно ограничиться случаем, когда оно равно трем. Это, как будет видно из дальнейшего, не отразится на общности рассуждений и результатов. На рис. 29 представлены эти три синусоиды в какой-то момент времени. Форма результирующего возмущения зависит от их взаимного расположения. Не нарушая общности, можно принять, что в этот момент какие-то три гребня синусоид A, B, C пространственно совпадают друг с другом.

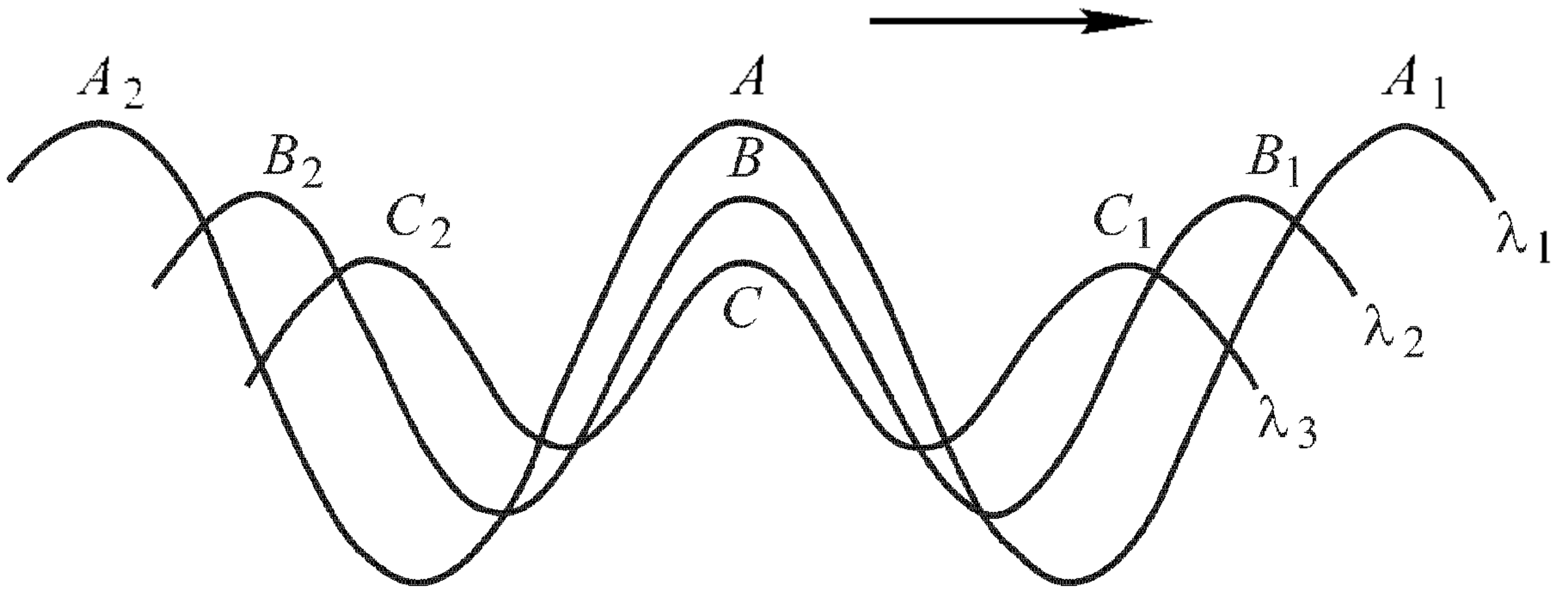


Рис. 29

Допустим ради определенности, что фазовая скорость v возрастает с возрастанием длины волны λ (если предположить противоположное, то рассуждения и окончательный результат не изменятся). Таким образом, если $\lambda_1 > \lambda_2 > \lambda_3$, а v_1, v_2, v_3 — соответствующие фазовые скорости, то $v_1 > v_2 > v_3$. При распространении, например, вправо более длинные волны будут обгонять более короткие, что приведет к непрерывному изменению формы результирующего возмущения. Гребни A, B, C начнут расходиться, гребни A_1, B_1, C_1 разойдутся еще больше, а гребни A_2, B_2, C_2 будут сближаться. Если начало координат поместить в точку, в которой находились гребни A, B, C в начальный момент, то координаты гребней A_2, B_2, C_2 в произвольный момент времени t представятся выражениями

$$x_{A_2}(t) = v_1 t - \lambda_1, \quad x_{B_2}(t) = v_2 t - \lambda_2, \quad x_{C_2}(t) = v_3 t - \lambda_3.$$

В момент времени τ, определяемый условиями

$$v_1\tau - \lambda_1 = v_2\tau - \lambda_2 = v_3\tau - \lambda_3,$$

т.е.

$$\tau = \frac{\lambda_1 - \lambda_2}{v_1 - v_2} = \frac{\lambda_2 - \lambda_3}{v_2 - v_3} = \frac{\lambda_1 - \lambda_3}{v_1 - v_3} = \frac{d\lambda}{dv} = \frac{1}{b},$$

гребни A_2, B_2, C_2 пространственно совпадут, так что первоначальное взаимное расположение синусоид, а с ним и форма всего возмущения восстановятся. Только роль гребней A, B, C перейдет к гребням A_2, B_2, C_2, что, очевидно, никак не может отразиться на форме результирующего возмущения. Полученный результат при линейном законе дисперсии (8.4) имеет общее значение, т.е. справедлив для произвольного числа синусоид и произвольного начального расположения их.

Таким образом, *по истечении времени $\tau = d\lambda/dv$, называемого временем восстановления, происходит периодическое восстановление формы возмущения.* Например, в рассмотренном нами случае трех синусоид в начале координат, где накладывались гребни A, B, C, в начальный момент времени был максимум возмущения. В точности такой же максимум появился через время τ в другом месте пространства, где наложились гребни A_2, B_2, C_2. Распространение возмущения носит как бы прыгающий характер, причем от прыжка к прыжку проходит время τ. Естественно определить скорость возмущения как отношение расстояния, проходимого возмущением за один «прыжок», к промежутку времени между последовательными прыжками. Так определенная величина называется *групповой скоростью* возмущения. В разобранном примере это будет отношение расстояния между двумя последовательными положениями максимума возмущения к времени восстановления τ. В момент времени $t = 0$ координата максимума $x_{\text{макс}}(0) = 0$. В момент τ координата такого же максимума будет

$$x_{\text{макс}}(\tau) = v_1\tau - \lambda_1 = v_2\tau - \lambda_2 = v_3\tau - \lambda_3 = v\tau - \lambda.$$

За время τ максимум проходит путь

$$x_{\text{макс}}(\tau) - x_{\text{макс}}(0) = v\tau - \lambda.$$

Следовательно, групповая скорость будет $u = v - \lambda/\tau$, или

$$u = v - \lambda\frac{dv}{d\lambda}. \tag{8.6}$$

Эта формула впервые была получена Рэлеем (1842–1919) и носит его имя. На рис. 30 приведена графическая интерпретация этой формулы, принадлежащая П.С. Эренфесту (1880–1933). На нем в координатах λ, v представлен график AB прямой (8.4). Так как $u = v - \lambda\, dv/d\lambda = a$, то эта прямая отсекает на оси ординат отрезок BO, длина которого равна групповой скорости u. Формулу (8.6) можно записать в виде

$$u = v + \frac{1}{\lambda}\frac{dv}{d(1/\lambda)} = \frac{d}{d(1/\lambda)}\left(\frac{v}{\lambda}\right),$$

или

$$u = \frac{d\omega}{dk}. \tag{8.7}$$

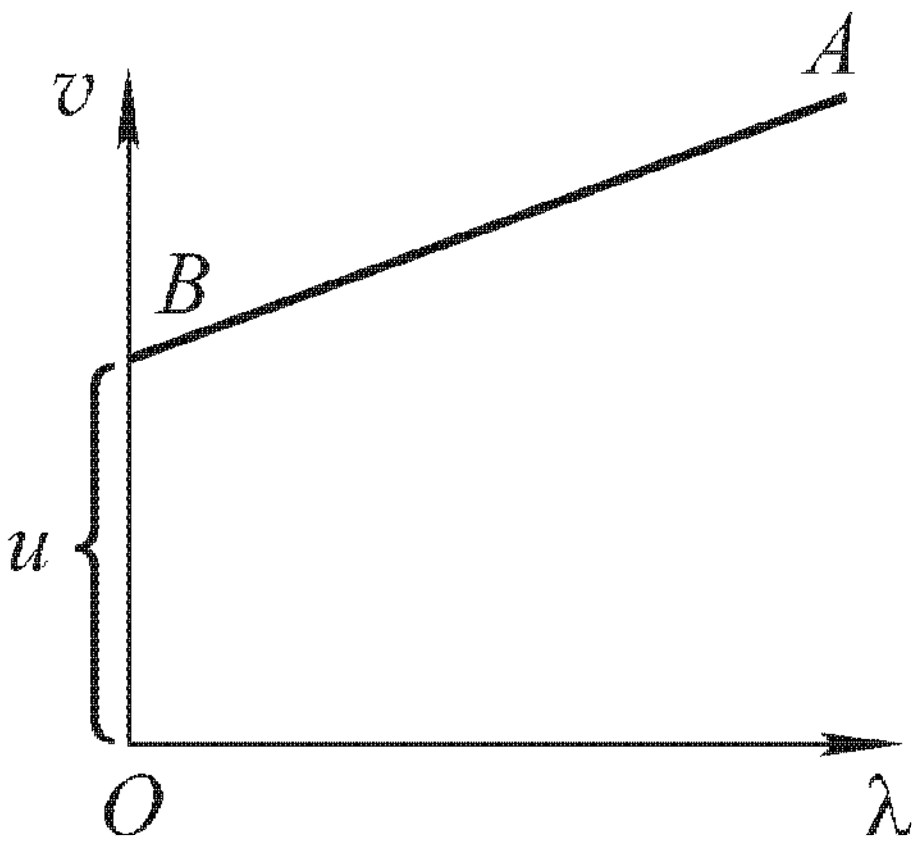


Рис. 30

Легко также преобразовать (8.6) к виду

$$u = \frac{c}{n}\left(1 + \frac{\lambda}{n}\frac{dn}{d\lambda}\right). \tag{8.8}$$

3. Полученные результаты строго справедливы при линейном законе дисперсии (8.4) (или (8.5)). Однако если возмущение занимает небольшую спектральную область, то эти результаты остаются приближенно верными и в диспергирующих непоглощающих средах. Возмущение такого типа называется *группой* *волн.* Точнее, *группой волн называется волновое образование, занимающее столь узкую спектральную область, что в пределах этой области приращение фазовой скорости v с достаточной точностью может считаться пропорциональным соответствующему приращению длины волны λ, а следовательно, приращение частоты ω — пропорциональным соответствующему приращению волнового числа k.* Это значит, что в пределах рассматриваемой спектральной области обе зависимости $v = v(\lambda)$ и $\omega = \omega(k)$ могут быть аппроксимированы линейными функциями λ и k, а именно

$$v = v(\lambda_0) + \left(\frac{dv}{d\lambda}\right)_{\lambda=\lambda_0} (\lambda - \lambda_0), \tag{8.9}$$

$$\omega = \omega(k_0) + \left(\frac{d\omega}{dk}\right)_{k=k_0} (k - k_0), \tag{8.10}$$

где λ_0 — какая-то длина волны, лежащая внутри спектральной области, занимаемой группой, а $k_0 = 2\pi/\lambda_0$ — соответствующее ей волновое число. При допустимости такой аппроксимации можно говорить и о времени приближенного восстановления формы возмущения $\tau = d\lambda/dv$, и о распространении возмущения с групповой скоростью u, определяемой выражением (8.6) или (8.7). На диаграмме Эренфеста в случае группы волн играет роль только малый участок кривой $u = v(\lambda)$, который можно приближенно считать прямолинейным и заменить соответствующим отрезком касательной. Групповая скорость u представится длиной отрезка OB, отсекаемого этой касательной на оси ординат (рис. 31).

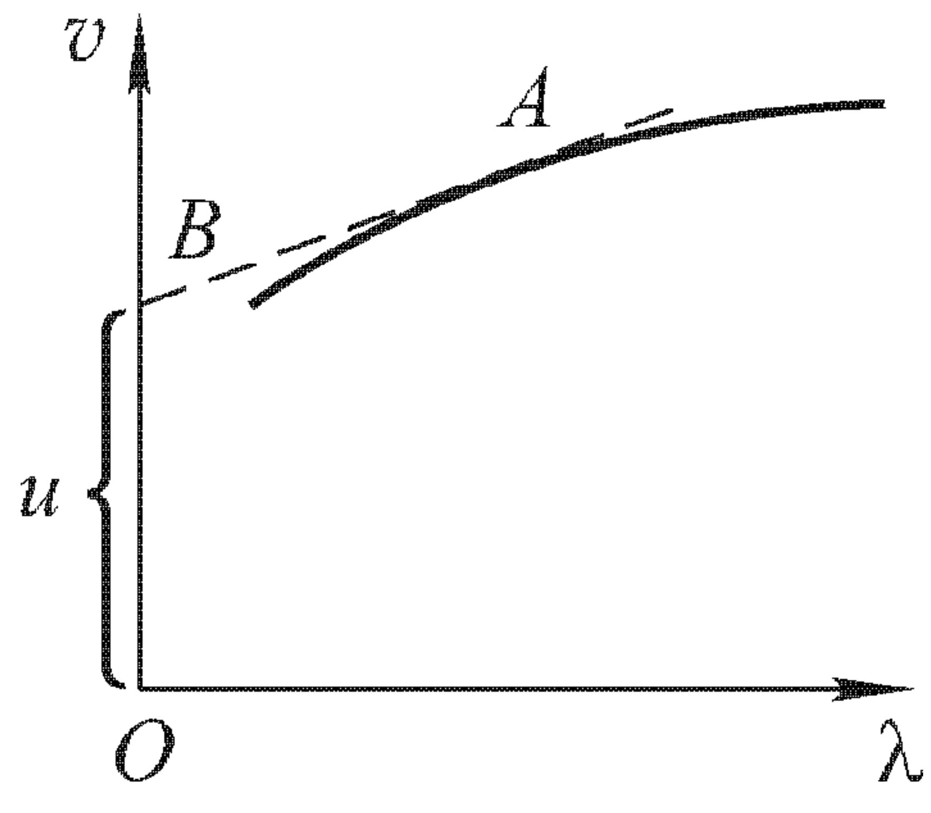


Рис. 31

4. Учет высших членов в разложении (8.9) или (8.10) приводит к следующему характеру распространения возмущения. Возмущение идет вперед, но его форма непрерывно изменяется. Однако по истечении времени $\tau = d\lambda/dv$ возмущение принимает форму, почти совпадающую с исходной, причем за это время оно продвигается вперед на расстояние $x = u\tau$. Можно сказать, что происходит передача энергии возмущения с групповой скоростью u. По истечении последующего промежутка времени той же длительности произойдет то же самое, и т.д. Вообще, в любой момент $t + \tau$ возмущение воспроизводит с малыми искажениями свою форму, какую оно имело в момент t, перемещаясь за время τ на расстояние $u\tau$. Но если возмущение распространяется достаточно долго, то малые изменения, претерпеваемые им за следующие друг за другом равные промежутки времени длительностью τ, будут накапливаться и могут настолько сильно исказить само возмущение, что его форма потеряет всякое сходство с исходной.

Чтобы оценить требующееся для этого время, дополним разложение (8.10) членом второй степени по $(k - k_0)$. С учетом формулы (8.7) можно написать:

$$\omega = \omega_0 + u_0(k - k_0) + \frac{1}{2}\left(\frac{du}{dk}\right)_0 (k - k_0)^2,$$

где нулем обозначены значения соответствующих величин при $k = k_0$. Пусть δk — максимальное значение разности $k - k_0$. Тогда соответствующее максимальное изменение фазы, обусловленное наличием квадратичного члена, будет $\frac{t}{2}\left|\frac{du}{dk}\right|(\delta k)^2$. Если это изменение мало по сравнению с величиной порядка π, то оно мало скажется на относительной разности фаз между синусоидами, входящими в группу, и тогда форма возмущения будет мало искажаться наличием квадратичного члена. Таким образом, чтобы на интервале времени t происходило периодическое восстановление исходной формы возмущения, необходимо выполнение условия

$$t \ll \frac{2\pi}{|du/dk|(\delta k)^2}. \tag{8.11}$$

Если перейти к длинам волн, то это условие преобразуется к виду

$$t \ll \frac{\lambda^2}{|du/d\lambda|(\delta\lambda)^2}. \tag{8.12}$$

Если же интервал времени t порядка или больше правой части этого неравенства, то о восстановлении исходной формы возмущения говорить не приходится.

5. Выше предполагалось, что плоские монохроматические волны, входящие в волновое образование, распространяются *в одном и том же направлении*. Рассмотрим теперь случай, когда такие волны занимают по-прежнему узкую область частот, но распространяются в разных направлениях, лежащих в пределах узкого конуса. Соответствующее волновое образование называется *волновым пакетом*. Его можно представить тройным интегралом

$$\mathbf{E}(\mathbf{r}, t) = \int\limits_{k_{0x} - \Delta k_x}^{k_{0x} + \Delta k_x} \int\limits_{k_{0y} - \Delta k_y}^{k_{0y} + \Delta k_y} \int\limits_{k_{0z} - \Delta k_z}^{k_{0z} + \Delta k_z} \mathbf{a}(\mathbf{k}) e^{i(\omega t - \mathbf{kr})}\, dk_x\, dk_y\, dk_z,$$

или сокращенно

$$\mathbf{E}(\mathbf{r}, t) = \int \mathbf{a}(\mathbf{k}) e^{i(\omega t - \mathbf{kr})}\, d\mathbf{k}. \tag{8.13}$$

Частоту ω следует рассматривать как функцию волнового вектора $\mathbf{k}$. Видом этой функции определяется *закон дисперсии* *волн.* Если среда *изотропна*, то функция $\omega(\mathbf{k})$ может зависеть только от *длины вектора* $\mathbf{k}$, но не от его направления. Но в *анизотропных средах*, например кристаллах, необходимо учитывать и зависимость ω *от направления* $\mathbf{k}$. Поэтому ниже вид функции $\omega(\mathbf{k})$ не конкретизируется, а рассуждения проводятся в общем виде. Полагая $\omega = \omega_0 + \Delta\omega$, $\mathbf{k} = \mathbf{k}_0 + \Delta\mathbf{k}$, аппроксимируем $\Delta\omega$ линейным выражением

$$\Delta\omega = \frac{\partial\omega}{\partial k_x}\Delta k_x + \frac{\partial\omega}{\partial k_y}\Delta k_y + \frac{\partial\omega}{\partial k_z}\Delta k_z = (\mathbf{u}\,\Delta\mathbf{k}), \tag{8.14}$$

где через $\mathbf{u}$ обозначен вектор с компонентами

$$u_x = \frac{\partial\omega}{\partial k_x}, \quad u_y = \frac{\partial\omega}{\partial k_y}, \quad u_z = \frac{\partial\omega}{\partial k_z}. \tag{8.15}$$

Сокращенно его записывают в символической форме:

$$\mathbf{u} = \frac{\partial\omega}{\partial\mathbf{k}}. \tag{8.16}$$

После подстановки соответствующих значений в выражение (8.13) оно преобразуется в

$$\mathbf{E} = \mathbf{A}(\mathbf{r}, t)e^{i(\omega_0 t - \mathbf{k}_0\mathbf{r})}, \tag{8.17}$$

где введено обозначение

$$\mathbf{A}(\mathbf{r}, t) = \int \mathbf{a}(\mathbf{k})e^{i(\Delta\omega t - \Delta\mathbf{k}\mathbf{r})}\,d\mathbf{k} = \int \mathbf{a}(\mathbf{k})e^{i(\mathbf{u}t - \mathbf{r})\Delta\mathbf{k}}\,d\mathbf{k}.$$

Отсюда видно, что в точке M, движущейся со скоростью $\mathbf{u}$ по закону $\mathbf{r} = \mathbf{u}t + \mathbf{r}_0$ ($\mathbf{r}_0 = \mathrm{const}$), амплитуда $\mathbf{A}$ остается постоянной. В такой точке совершаются гармонические колебания

$$\mathbf{E} = \mathbf{A}e^{i(\omega_0 t - \mathbf{k}_0\mathbf{r})} = \mathbf{A}e^{i[(\omega_0 - \mathbf{k}_0\mathbf{u})t - \mathbf{k}_0\mathbf{r}_0]}$$

с частотой $\omega_0 - \mathbf{k}_0\mathbf{u}$. По истечении времени τ фаза этих колебаний изменяется на $(\omega_0 - \mathbf{k}_0\mathbf{u})\tau$. Если это изменение равно 2π, т.е.

$$\tau = \frac{2\pi}{\omega_0 - \mathbf{k}_0\mathbf{u}} \approx \frac{2\pi}{\omega - \mathbf{k}\mathbf{u}}, \tag{8.18}$$

то вектор $\mathbf{E}$ в движущейся точке M в любой момент времени t примет то же значение, какое он имел в более ранний момент $t - \tau$. Так как это справедливо при любом значении параметра $\mathbf{r}_0$, то происходит периодическое восстановление формы возмущения с периодом τ, причем за это время возмущение перемещается вперед на расстояние $\mathbf{u}\tau$. В результате мы снова приходим к представлению о распространении возмущения с групповой скоростью $\mathbf{u}$, определяемой выражением (8.16).

В изотропных средах векторы $\mathbf{u}$ и $\mathbf{k}$ параллельны. В этом случае (8.18) легко преобразовать к прежнему виду $\tau = d\lambda/dv$. Однако в анизотропных средах векторы $\mathbf{u}$ и $\mathbf{k}$, вообще говоря, не параллельны, и надо пользоваться более общими выражениями (8.16) и (8.18). Конечно, и здесь из-за наличия членов высших степеней, отброшенных в разложении (8.14), за время τ происходит не точное, а лишь *приближенное* восстановление формы возмущения. За это время возмущение претерпевает малые, может быть едва заметные, искажения. Но на больших интервалах времени эти искажения накапливаются и исходная форма возмущения сможет претерпеть существенные изменения.

6. Изложенные соображения позволяют легко решить вопрос о *скорости движения энергии* или *светового сигнала* в диспергирующих средах в тех областях спектра, в которых применимо понятие групповой скорости (т.е. вдали от полос поглощения). Прежде всего заметим, что фазовая скорость не имеет ничего общего со скоростью движения энергии. Фазовой скоростью устанавливается только связь *между фазами колебаний* в различных точках пространства. Связь такого типа в принципе может существовать и без передачи энергии, как это видно из следующего примера.

Вообразим длинную цепочку спортсменов, расположенных вдоль прямой линии на равных расстояниях друг от друга. Пусть они выполняют одно и то же гимнастическое упражнение, например периодическое движение руками, и притом так, что каждый впереди стоящий спортсмен начинает движение с некоторым запаздыванием по отношению к спортсмену, стоящему за ним. Пусть время запаздывания одно и то же для всех спортсменов. При наблюдении со стороны будет казаться, что по цепочке бежит волна с определенной фазовой скоростью, значение которой зависит от расстояния между соседними спортсменами и от времени запаздывания, о котором говорилось выше. Наличие такой волны, конечно, не означает, что каждый спортсмен приводит в движение впереди стоящего спортсмена. Так и возможность распространения в среде плоской монохроматической волны еще не дает оснований для заключения о переносе энергии с фазовой скоростью.

Строго плоская монохроматическая волна непригодна для наблюдения передачи энергии, поскольку она не имеет ни начала, ни конца во времени и в пространстве. Сама постановка вопроса о передаче энергии требует отказа от такой идеализации. Необходимо перейти к волновому возмущению, *ограниченному в пространстве* по крайней мере с одного конца, т.е. имеющему передовой фронт, перед которым возмущение отсутствует. Подходящим волновым образованием может служить группа волн. Если выполнено условие (8.12), то средняя скорость энергии, переносимой группой, совпадает с групповой скоростью. Действительно, форма группы, какую она имела в момент t, восстанавливается без заметного искажения в более поздний момент времени $t+\tau$. При этом группа вместе с локализованной в ней энергией за время τ перемещается вперед на расстояние $x = u\tau$. Так как такое восстановление формы имеет место, каков бы ни был момент времени t, то движение энергии с групповой скоростью будет происходить на протяжении как угодно длинного промежутка времени, даже если за это время группа существенно изменит свою форму.

Итак, в области, далекой от области сильного поглощения, скорость движения энергии в группе волн совпадает с групповой скоростью. То же самое приближенно справедливо и для скорости движения энергии в волновом возмущении, занимающем сравнительно широкую спектральную область, если только в пределах этой спектральной области групповая скорость $u = u(\lambda)$ меняется мало. Если ширина спектральной области $\delta\lambda$, занимаемой группой, стремится к нулю, то группа в пределе переходит в монохроматическую волну. Можно поэтому сказать, что средняя скорость переноса энергии в монохроматической волне совпадает с групповой скоростью. Это утверждение следует понимать именно в приведенном смысле, рассматривая монохроматическую волну как предельный случай квазимонохроматической. Нельзя ограничиться идеализированной плоской строго монохроматической волной, отвлекаясь от представления ее как предельного случая квазимонохроматической волны. При такой абстрактной постановке вопроса утрачивается связь с реальными явлениями, а потому с точки зрения физики она бессмысленна.

Прямые измерения скорости света сводятся к измерению расстояния, проходимого световым сигналом за определенный промежуток времени. Из изложенного выше следует, что этот метод практически дает групповую скорость. То же самое, как показывает подробный анализ, относится ко всем известным косвенным методам измерения скорости света. Фазовую скорость, точнее — отношение фазовых скоростей в двух различных средах, можно определить по отношению показателей преломления, используя формулу волновой теории (3.7), в которую входят фазовые скорости света в рассматриваемых средах (см. § 64).

7. Остановимся еще на вопросе о *скорости распространения передового фронта волнового возмущения*. Речь идет о волне, резко ограниченной передовым фронтом, перед которым никакого возмущения нет. Скорость такого фронта точно совпадает со скоростью света в вакууме c. В этом легко убедиться, исходя из основных представлений электронной теории. Согласно этой теории, всякую среду следует рассматривать как вакуум, в который вкраплены молекулы и атомы вещества. Свет распространяется в вакууме между атомами и молекулами вещества, т.е. всегда со скоростью c. Когда световое возмущение достигает какого-либо атома, электроны и атомные ядра приходят в колебания и сами становятся центрами излучения новых электромагнитных волн. Эти вторичные волны накладываются на первичную волну и тем самым определяют все волновое поле в среде. Но из-за инерции электроны и ядра не сразу приходят в колебания. Пока электроны и ядра не пришли в колебания, они не излучают вторичные волны, а потому не оказывают влияния на распространение возмущения. Поэтому ясно, что передовой фронт должен распространяться в среде с той же скоростью, что и в вакууме. Но почему же при измерении скорости света получается не c, а другая величина? Дело в том, что передовой фронт несет слишком малую энергию, а приемники света недостаточно чувствительны, чтобы ее обнаружить. Количественные расчеты, выполненные впервые Зоммерфельдом (1868–1951) и более подробно Л. Бриллюэном (1889–1969), показали, что это действительно так.

ГЛАВА II

ГЕОМЕТРИЧЕСКАЯ ТЕОРИЯ ОПТИЧЕСКИХ ИЗОБРАЖЕНИЙ

§ 9. Понятие оптического изображения

1. Если пучок световых лучей, исходящий из какой-либо точки P, в результате отражений, преломлений или изгибаний в неоднородной среде сходится в точке P', то P' называется *оптическим изображением* или просто *изображением точки* P. Точку P' называют также *фокусом геометрического схождения лучей*. Изображение P' называется *действительным*, если световые лучи действительно пересекаются в точке P'. Если же в P' пересекаются продолжения лучей, проведенные в направлении, обратном распространению света, то изображение называется *мнимым*. При помощи оптических приспособлений мнимые изображения могут быть преобразованы в действительные. Например, в нашем глазу мнимое изображение преобразуется в действительное, получающееся на сетчатке глаза.

Если в некоторый момент времени изменить на противоположное направление магнитного или электрического вектора, то, согласно *принципу обратимости* (см. т. III, § 83, п. 8), форма лучей остается без изменения, но направление распространения света изменится на противоположное. Точка P' будет играть роль *источника света*, а P — *его изображения*. Поэтому P и P' называются *сопряженными* или *взаимно сопряженными точками*. Аналогично, две линии или две поверхности называются *сопряженными*, если одна из них является оптическим изображением другой.

Если желают подчеркнуть, что лучи строго пересекаются в точке P', то изображение называют *стигматическим*. Пучок же лучей, исходящих из одной точки или сходящихся в одной точке, называется *гомоцентрическим*. Примером может служить отражение от *эллипсоидального зеркала*. По свойству эллипсоида вращения прямые FA и $F'A$ (рис. 32), соединяющие его фокусы F и F' с произвольной точкой A поверхности эллипсоида, наклонены под одинаковыми углами к этой поверхности. Поэтому все лучи, вышедшие из одного фокуса, после отражения от поверхности эллипсоида пересекутся в другом фокусе. Практически более важен случай *параболоидального зеркала*, используемого в астрономических телескопах-рефлекторах. Параболоидальное зеркало есть частный случай эллипсоидального, когда один из фокусов F' удален в бесконечность. В силу известного свойства параболы все лучи, параллельные оси параболоида, после отражения от его вогнутой поверхности пересекутся в фокусе параболоида F (рис. 33). Если же такие лучи отразятся от выпуклой поверхности параболоида, то в фокусе F пересекутся продолжения отраженных лучей (рис. 34). Стигматическое изображение точки P в виде точки P' получается также при преломлении на поверхности, являющейся *анаберрационной* для этой пары точек (см. § 7 и задачи к настоящему параграфу).

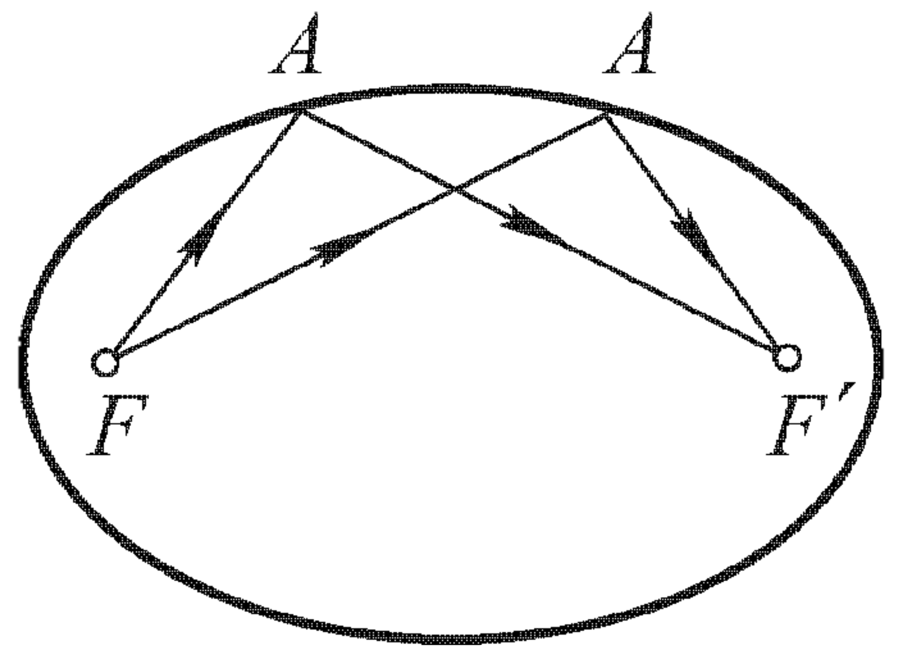


Рис. 32

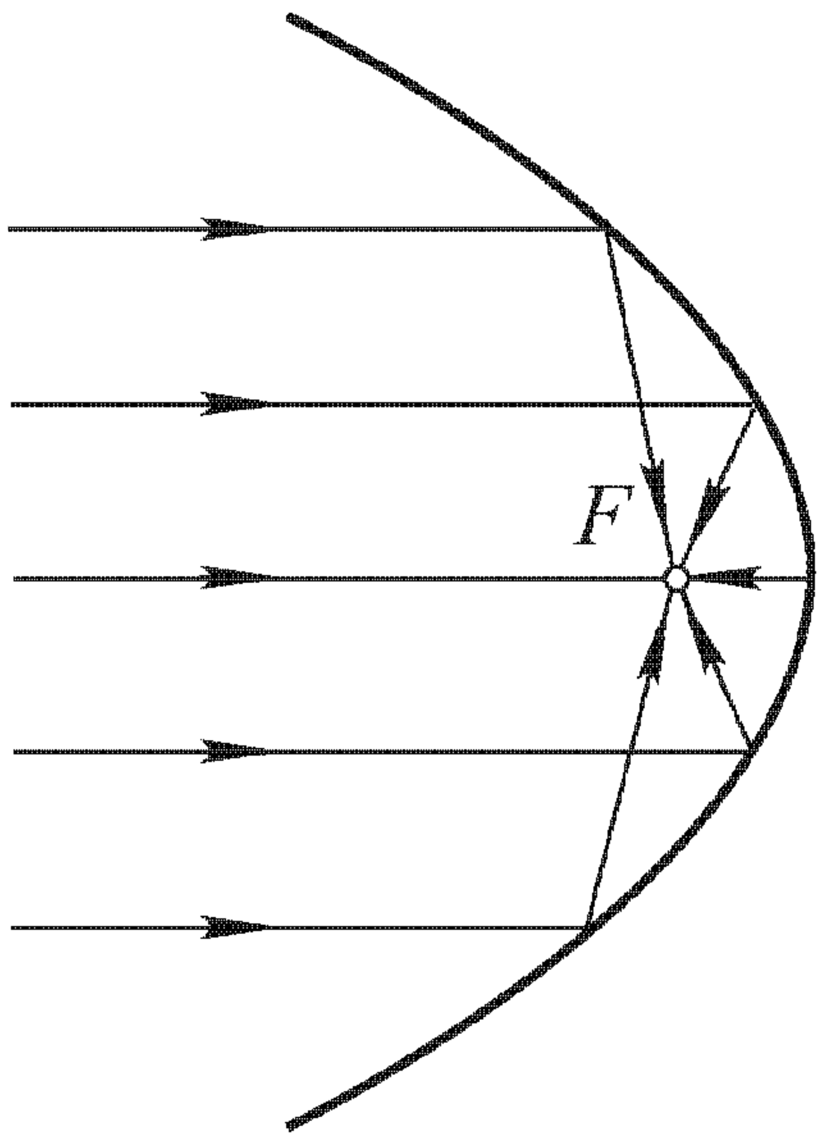


Рис. 33

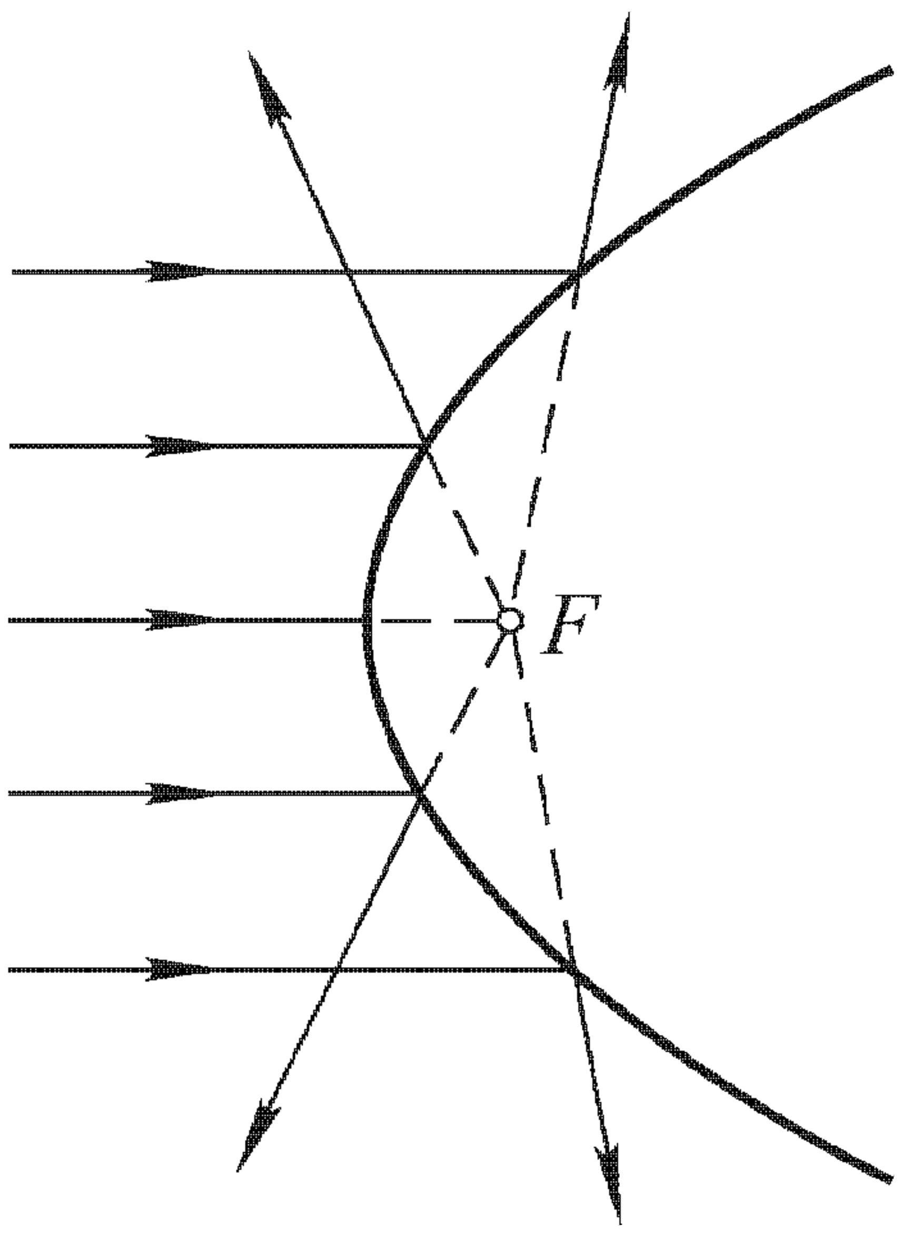


Рис. 34

На практике случаи стигматических изображений, как правило, бывают исключениями. Обычно лучи пересекаются не строго в одной точке, а в некоторой окрестности ее. Изображением светящейся точки на экране будет в этих случаях не математическая точка, а *светлое пятнышко*. Это снижает качество изображения. Однако строго точечное изображение светящейся точки не получается даже в тех случаях, когда по законам геометрической оптики лучи должны точно пересекаться в одной точке. Из-за дифракции света изображение светящейся точки получается в виде светлого кружка, окруженного темными и светлыми кольцами.

Полная — *физическая* — теория оптических изображений должна учитывать *волновые свойства света*. Но начинать с такой теории, из-за ее сложности, было бы нецелесообразно. Сначала надо уяснить получение изображений с простейшей — *геометрической* — точки зрения, а затем ввести поправки, учитывающие волновую природу света. Геометрическая теория строится на основе одних только законов отражения и преломления света, полностью отвлекаясь от его физической природы. Таким путем, конечно, нельзя установить границы применимости геометрической теории. Это можно сделать только на основе волновой теории. Но значение последней этим не исчерпывается. Более существенно, что волновая теория позволяет определить, *чтó принципиально возможно в оптике*, и указать, как этого достигнуть.

2. Непрерывная совокупность точек, изображаемых оптической системой, называется *пространством предметов*. Непрерывная же совокупность точек, являющихся их изображениями, называется *пространством изображений*. Одна и та же точка может принадлежать к пространству предметов и к пространству изображений, в зависимости от того, рассматривается ли она как *предмет* или как *изображение*. Показатель преломления пространства предметов будем обозначать через n, а пространства изображений — через n'. В оптических приборах световой оптики величины n и n' всегда постоянны, т.е. не меняются от точки к точке. Когда изображение действительное, под n' следует понимать показатель преломления среды *в той точке, где получилось это изображение*. В случае мнимого изображения n' не всегда совпадает с показателем преломления среды в месте нахождения точки P'. Значение n' относится *к той среде, через которую проходят действительные лучи, продолжения которых пересекаются в точке* P'.

Мы будем исследовать и такие случаи, когда показатель преломления среды меняется *непрерывно* от точки к точке, а потому лучи *криволинейны*. Такой случай практически реализуется в *электронной оптике*. Здесь роль линз выполняют *электрические и магнитные поля*, а показателя преломления — *скорость электрона* (см. § 4).

3. С математической точки зрения задача геометрической теории оптических изображений сводится к определению положения изображения при любом заданном положении предмета. При этом общие свойства оптических систем удобно исследовать с помощью следующего положения. *Оптические длины всех лучей, соединяющих сопряженные точки* P *и* P', *одинаковы*. Это непосредственно очевидно, когда изображение P' действительное, так как тогда сферическая волна, вышедшая из P, превращается в сферическую волну, сходящуюся в P'. *Оптические же длины всех лучей от одного положения волнового фронта до другого одинаковы.* Но это положение можно распространить и на *мнимые изображения.* В этом случае не существует лучей, соединяющих P с P'. Роль луча играет прямолинейное продолжение его в сторону изображения P'. По аналогии с мнимым изображением такое продолжение можно назвать *мнимым лучом.*

Оптическую длину луча следует считать *положительной*, когда он проходится в направлении распространения света, и *отрицательной* в противоположном случае. Чтобы в случае мнимых изображений избежать неопределенности, будем предполагать, что пространство изображений *однородно*, т.е. световые лучи в нем *прямолинейны.* Это не значит, что изображение P' должно обязательно получаться в том месте, где среда однородна. Требуется только, чтобы действительные световые лучи, продолжения которых сходятся в P', были *прямолинейны.*

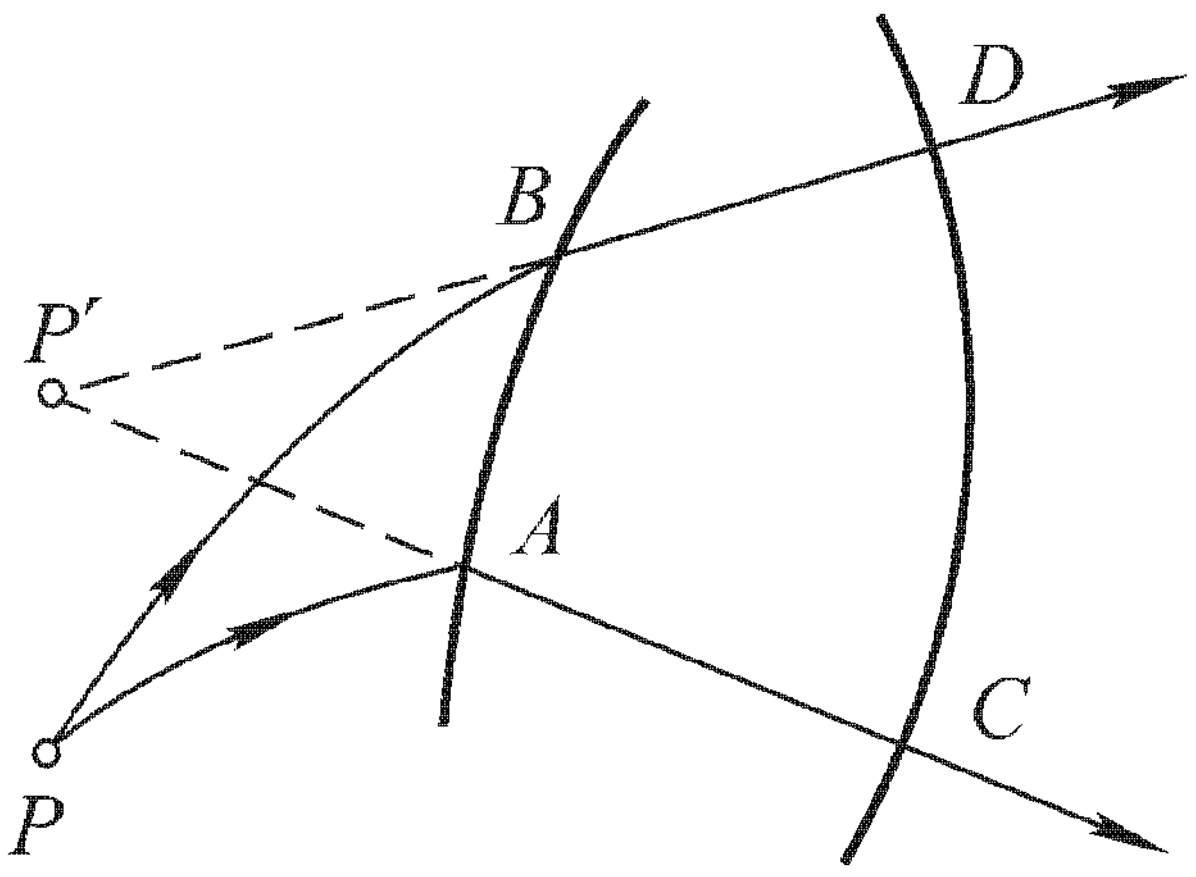


Рис. 35

После этих замечаний обратимся к доказательству нашего утверждения. Пусть лучи PAC и PBD (рис. 35), вышедшие из точки P, на участках AC и BD прямолинейны. Их продолжения пересекаются в точке P', являющейся мнимым изображением точки P. Волновой фронт в однородном пространстве изображений будет иметь форму сферы CD с центром в P'. Очевидно,

$$(PAC) = (PBD), \quad (P'AC) = (P'BD).$$

Почленное вычитание дает

$$(PA) - (P'A) = (PB) - (P'B).$$

Но, согласно нашему правилу знаков,

$$(PA) - (P'A) = (PA) + (AP') = (PAP'),$$

$$(PB) - (P'B) = (PB) + (BP') = (PBP').$$

Следовательно, $(PAP') = (PBP')$, что и требовалось доказать.

Доказанное свойство оптических длин эквивалентно утверждению, что *свет затрачивает одно и то же время, распространяясь вдоль различных лучей от точечного источника до его изображения.* В таком виде это утверждение называется *принципом таутохронизма* (равенства времен распространения). Принципом таутохронизма мы воспользуемся при изучении явлений интерференции.

Наряду с мнимыми изображениями, следует ввести и *мнимые источники света*, или *мнимые объекты*. Точечный объект называется *мнимым*, если он является точкой пересечения продолжений действительных лучей, проведенных в обратных направлениях. Мнимый объект можно рассматривать как источник *мнимых лучей*. Из множества точечных мнимых объектов составляются мнимые объекты конечных размеров.

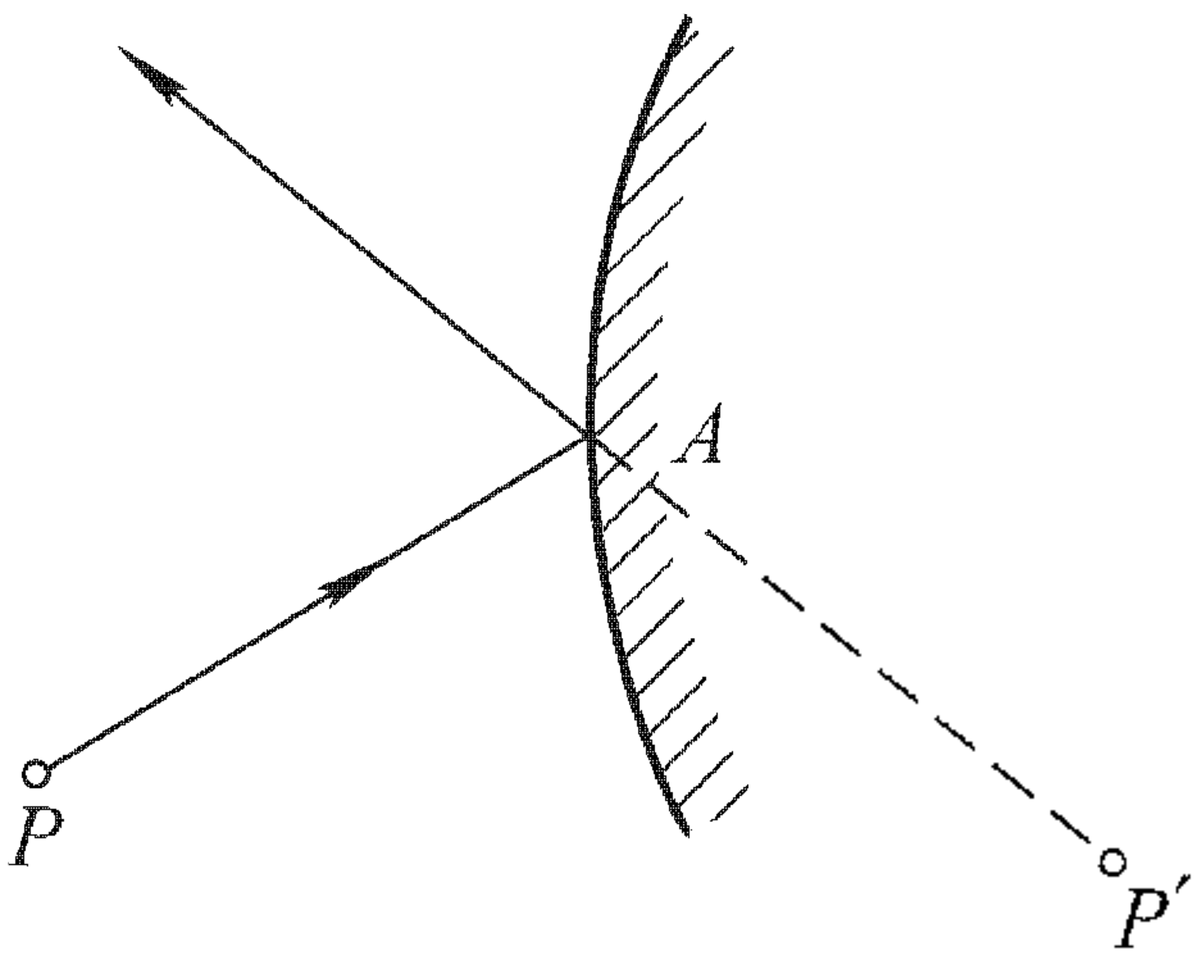


Рис. 36

Введение мнимых объектов и мнимых лучей освобождает теорию от необходимости раздельного рассмотрения действительных и мнимых изображений. Отпадает также необходимость в раздельном рассмотрении преломления и отражения света, что имеет большое значение в теории оптических систем, содержащих большое количество преломляющих и отражающих поверхностей. Действительно, пусть P' — мнимое изображение точки P, полученное в результате отражения света от зеркала (рис. 36). Согласно принятому нами правилу знаков, оптическая длина мнимого луча AP' отрицательна. Поэтому для оптической длины пути PAP' можно написать:

$$(PAP') = n|PA| - n|AP'| = n|PA| + n'|AP'|,$$

где $n' = -n$. Поэтому *отражение формально математически можно рассматривать как преломление, если только показателю преломления n' приписать отрицательное значение* $(n'/n = -1)$.

ЗАДАЧИ

1. Две однородные среды с показателями преломления n и n' граничат друг с другом вдоль поверхности S (рис. 37), являющейся поверхностью вращения вокруг оси OP' (оптической оси). Найти форму поверхности S, при которой она будет анаберрационной для пары точек P и P', лежащих на оптической оси, из которых точка P удалена в бесконечность, а P' может занимать любое положение на оптической оси.

Решение. Примем оптическую ось за координатную ось X, начало координат поместим в точке пересечения ее с поверхностью S, ось Y направим вверх перпендикулярно к оптической оси. Так как оптические длины лучей от бесконечно удаленной точки P до плоскости OA одинаковы, то условие анаберрационности поверхности S будет $(ABP') = (OP')$, или

$$nx + n'\sqrt{(x-q)^2 + y^2} = n'q,$$

где x и y — текущие координаты точек поверхности S, а q — абсцисса точки P'. Перенеся nx в правую часть и возведя в квадрат, находим уравнение искомой поверхности:

$$(n'^2 - n^2)x^2 + n'^2y^2 - 2n'(n' - n)qx = 0. \tag{9.1}$$

Допустим сначала, что $n'^2 - n^2 > 0$. Тогда уравнение (9.1) представляет эллипсоид вращения с полуосями

$$a = \frac{n'}{n' + n}q, \quad b = \sqrt{\frac{n' - n}{n' + n}}\,q.$$

Эллипсоид вытянут в направлении оси X. Его эксцентриситет равен

$$e = \frac{\sqrt{a^2 - b^2}}{a} = \frac{n}{n'} < 1.$$

Изображение P' действительное.

Пусть теперь $n'^2 - n^2 < 0$. Тогда (9.1) есть уравнение двуполостного гиперболоида вращения с полуосями

$$a = \frac{n'}{n' + n}q, \quad b = \sqrt{\frac{n - n'}{n' + n}}\,q$$

и эксцентриситетом

$$e = \frac{\sqrt{a^2 + b^2}}{a} = \frac{n}{n'} > 1.$$

Изображение P' мнимое (рис. 38).

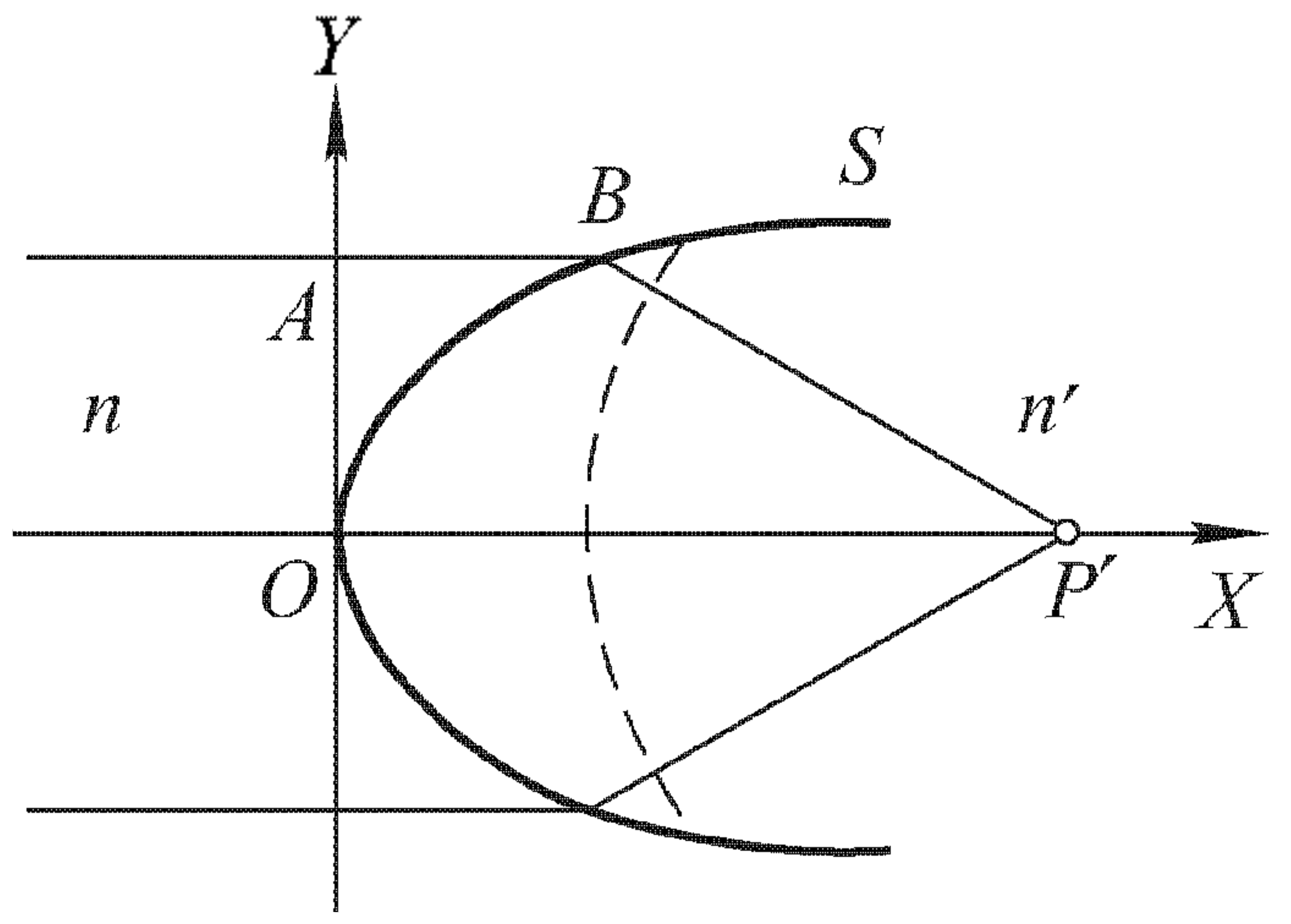


Рис. 37

Рассмотрим, наконец, случай, когда $n'^2 - n^2 = 0$. Это может быть либо при $n' - n = 0$, либо при $n' + n = 0$. Первая возможность не представляет интереса, поскольку она соответствует тривиальному случаю, когда обе граничащие среды в оптическом отношении тождественны. Вторая возможность $n' = -n$ может быть реализована при отражении света. В этом случае уравнение (9.1) переходит в

$$y^2 = 4qx \tag{9.2}$$

и представляет параболоид вращения с параметром $p = 2q$ (параболоидальное зеркало). Если $q > 0$ (см. рис. 34), то фокус P' (на рис. 34 — точка F) мнимый. Если $q < 0$ (см. рис. 33), то он действительный. Эти случаи уже были рассмотрены в тексте.

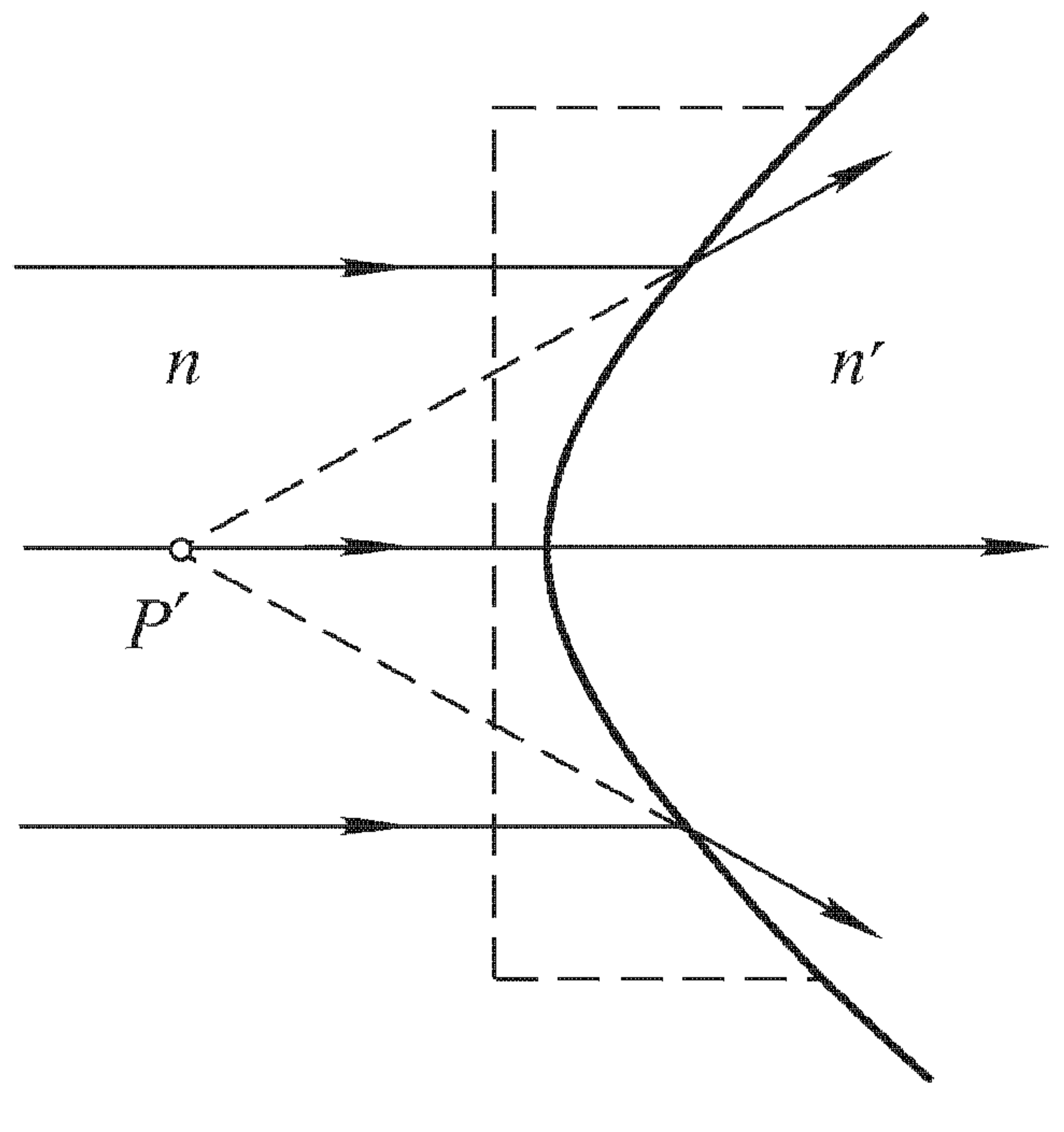


Рис. 38

Результаты решения этой задачи указывают способ построения *идеальной линзы* для пары сопряженных точек, из которых одна бесконечно удаленная. Рассмотрим сначала линзу, ограниченную поверхностью эллипсоида вращения OB и сферической поверхностью с центром в P' (на рис. 37 эта поверхность изображена штриховой линией). Эксцентриситет эллипсоида должен быть равен $1/n$, где n — показатель преломления линзы относительно окружающей среды. Параллельный пучок лучей, падая на поверхность эллипсоида, после преломления на ней превращается в пучок, сходящийся в точке P'. Задняя — сферическая — поверхность линзы не меняет направления лучей, поскольку ее центр находится в точке схождения пучка P'. Таким образом, рассматриваемая линза собирает параллельный пучок лучей строго в одной точке P'. Если точечный источник поместить в P', то после прохождения через линзу пучок лучей станет строго параллельным оптической оси.

Рассмотрим далее линзу, ограниченную плоской поверхностью (на рис. 38 она изображена штриховой линией) и гиперболоидом вращения с эксцентриситетом n. Параллельный пучок лучей, падающих на плоскую поверхность линзы, после прохождения через эту поверхность не изменит направления, а после преломления на поверхности гиперболоида превратится в расходящийся пучок лучей, продолжения которых пересекаются строго в одной точке P'.

2. Найти уравнение картезианского овала (см. § 7, п. 6).

Решение. Пусть $P(q, 0)$ и $P'(q', 0)$ — сопряженные точки, для которых поверхность, получающаяся от вращения картезианского овала относительно оси симметрии PP', анаберрационна. Поместим начало координат в точку перечения овала с прямой PP'. Тогда по определению анаберрационной поверхности

$$n\sqrt{(x-q)^2+y^2}+n'\sqrt{(x-q')^2+y^2}=n'q'-nq.$$

Освобождаясь от радикалов, получим уравнение картезианского овала:

$$\begin{aligned}(n^2-n'^2)(x^2+y^2)+4(n^2-n'^2)(n'^2q'-n^2q)(x^2+y^2)x+\\+4nn'(nq-n'q')(nq'-n'q)(x^2+y^2)+4(n'^2q'-n^2q)^2x^2+\\+8nn'(n'-n)(nq-n'q')qq'x=0.\end{aligned}\tag{9.3}$$

При определенных значениях параметров n, n', q, q' картезианской овал вырождается в поверхности второго порядка. Тогда получаются, в частности, уже разобранные ранее случаи, изображенные на рисунках 32, 33, 34, 37 и 38.

§ 10. Преломление на сферической поверхности. Сферические зеркала и тонкие линзы

1. Важнейшие из оптических инструментов или их составные части относятся к так называемым *центрированным оптическим системам*. Они представляют собой оптически однородные преломляющие или отражающие среды, отделенные одна от другой сферическими поверхностями, центры кривизны которых расположены на одной прямой, называемой *главной оптической осью* системы. Обычно если это не может привести к недоразумениям, прилагательное «главная» мы будем опускать.

2. Начнем с простейшего случая одной сферической преломляющей поверхности, разграничивающей однородные среды с показателями преломления n и n'. Можно предполагать (хотя это и не обязательно), что эта поверхность обладает *симметрией вращения* относительно одной из прямых OC, проходящих через *центр кривизны* сферической поверхности (рис. 39). Такая прямая и будет главной оптической осью. Примем ее за координатную ось X. Начало координат поместим в точке O, в которой главная оптическая ось пересекает сферическую поверхность.

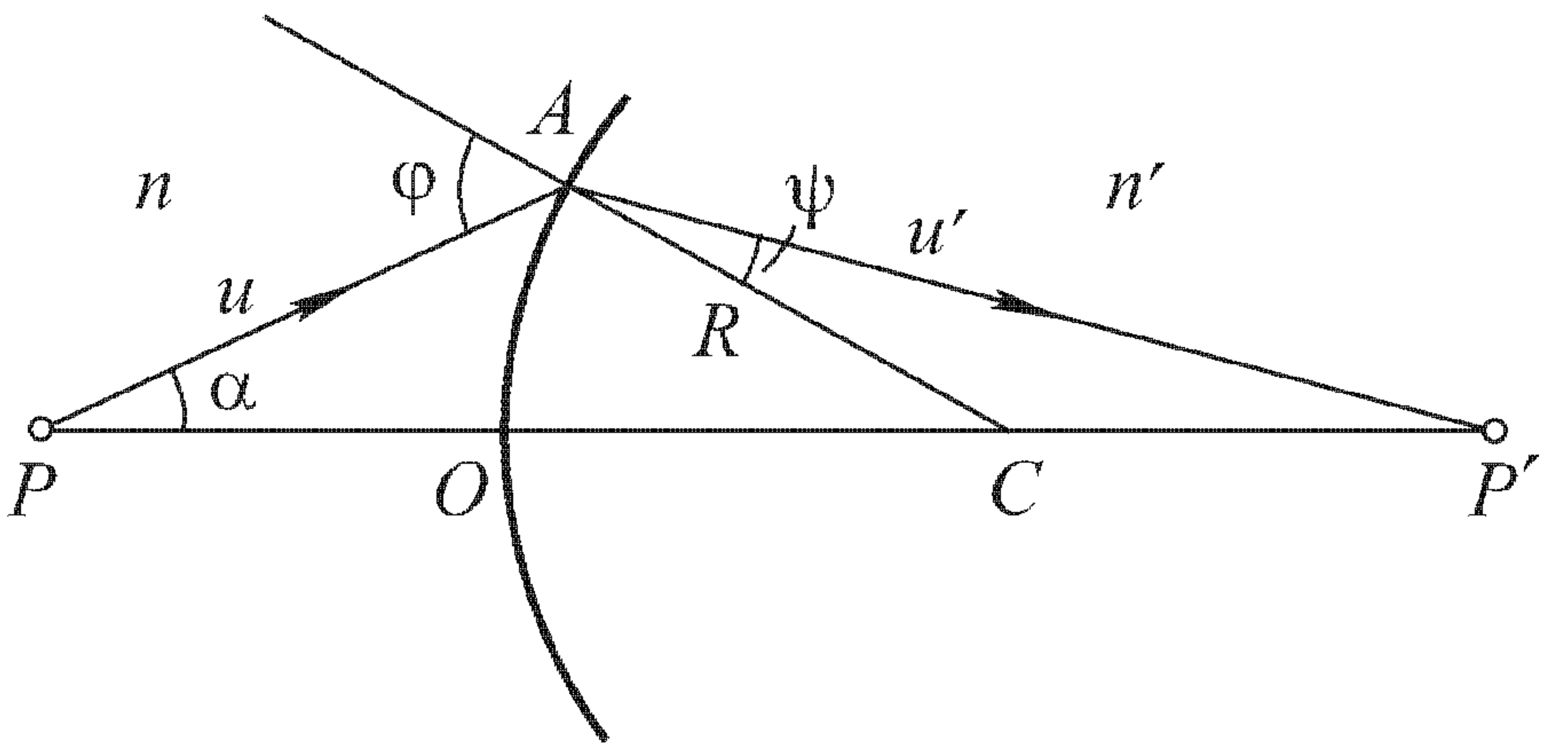


Рис. 39

Ввиду симметрии вращения достаточно ограничиться рассмотрением хода лучей в координатной плоскости XY. Совместим ее с плоскостью рисунка. Абсциссы и ординаты будем отсчитывать от начала координат O. Если направление отсчета совпадает с направлением распространения света вдоль оптической оси, то соответствующая абсцисса считается *положительной*; в противоположном случае она считается *отрицательной*. То же относится и ко всем другим направленным отрезкам. Например, на рис. 39 абсцисса точки P отрицательна, а точка P' положительна. Ордината считается положительной, если соответствующая точка лежит выше оптической оси, и отрицательной, когда она расположена ниже.

Допустим, что точечный источник света P находится на оптической оси системы (рис. 39). Произвольный луч PA после преломления на сферической поверхности пойдет по пути AP'. Обозначим длины AP и AP' через u и u' соответственно. Эти длины отсчитываются от точки A и считаются положительными, если направление отсчета совпадает с направлением распространения света, и отрицательными в противоположном случае. Из рисунка видно, что

$$\text{пл. } PAC + \text{пл. } CAP' = \text{пл. } PAP'.$$

Так как $u < 0$, $u' > 0$, то для этих площадей можно написать:

$$\text{пл. } PAC = \frac{1}{2}\,|PA|\cdot|AC|\sin\varphi = -\frac{1}{2}\,uR\sin\varphi,$$

$$\text{пл.} CAP' = \frac{1}{2}\,u'R\sin\psi,$$

$$\text{пл.} PAP' = -\frac{1}{2}\,uu'\sin(\varphi-\psi) = -\frac{1}{2}\,uu'(\sin\varphi\cos\psi - \sin\psi\cos\varphi).$$

Здесь R — радиус кривизны преломляющей поверхности. Он отсчитывается от сферической поверхности к ее центру (на рис. 39

радиус R положителен). Таким образом,

$$-uR\sin\varphi + u'R\sin\psi = -uu'(\sin\varphi\cos\psi - \sin\psi\cos\varphi).$$

По закону преломления $\sin\varphi/\sin\psi = n'/n$, а потому

$$-u'Rn' + u'Rn = uu'(n\cos\varphi - n'\cos\psi).$$

Отсюда

$$\frac{n}{u} - \frac{n'}{u'} = \frac{n\cos\varphi - n'\cos\psi}{R}. \tag{10.1}$$

Положение точки P' зависит от угла наклона α падающего луча к оптической оси. Ограничимся, однако, малыми углами α и допустим, что углы φ и ψ также малы. Лучи, удовлетворяющие таким условиям, называются *параксиальными* (*приосевыми*) *лучами*. Для них можно принять

$$\cos\varphi = \cos\psi = 1, \quad AP \approx OP = x, \quad AP' \approx OP' = x'.$$

В этом приближении формула (10.1) переходит в

$$\frac{n}{x} - \frac{n'}{x'} = \frac{n-n'}{R}. \tag{10.2}$$

Отсюда видно, что в рассматриваемом приближении положение точки P' не зависит от угла α. Следовательно, все параксиальные лучи, выходящие из одной точки оптической оси, после преломления на сферической поверхности пересекутся приближенно в одной точке, лежащей также на оптической оси. Точка P' будет поэтому *оптическим изображением точки P в параксиальных лучах.* Во всем дальнейшем предполагается, что все лучи, проходящие через центрированные системы, параксиальны.

При выводе предполагалось, что источник света P действительный. Однако все сказанное справедливо и для мнимого источника, т.е. тогда, когда источником служит точка схождения продолжений падающих лучей. В этом легко убедиться, повторив рассуждения применительно к рис. 40.

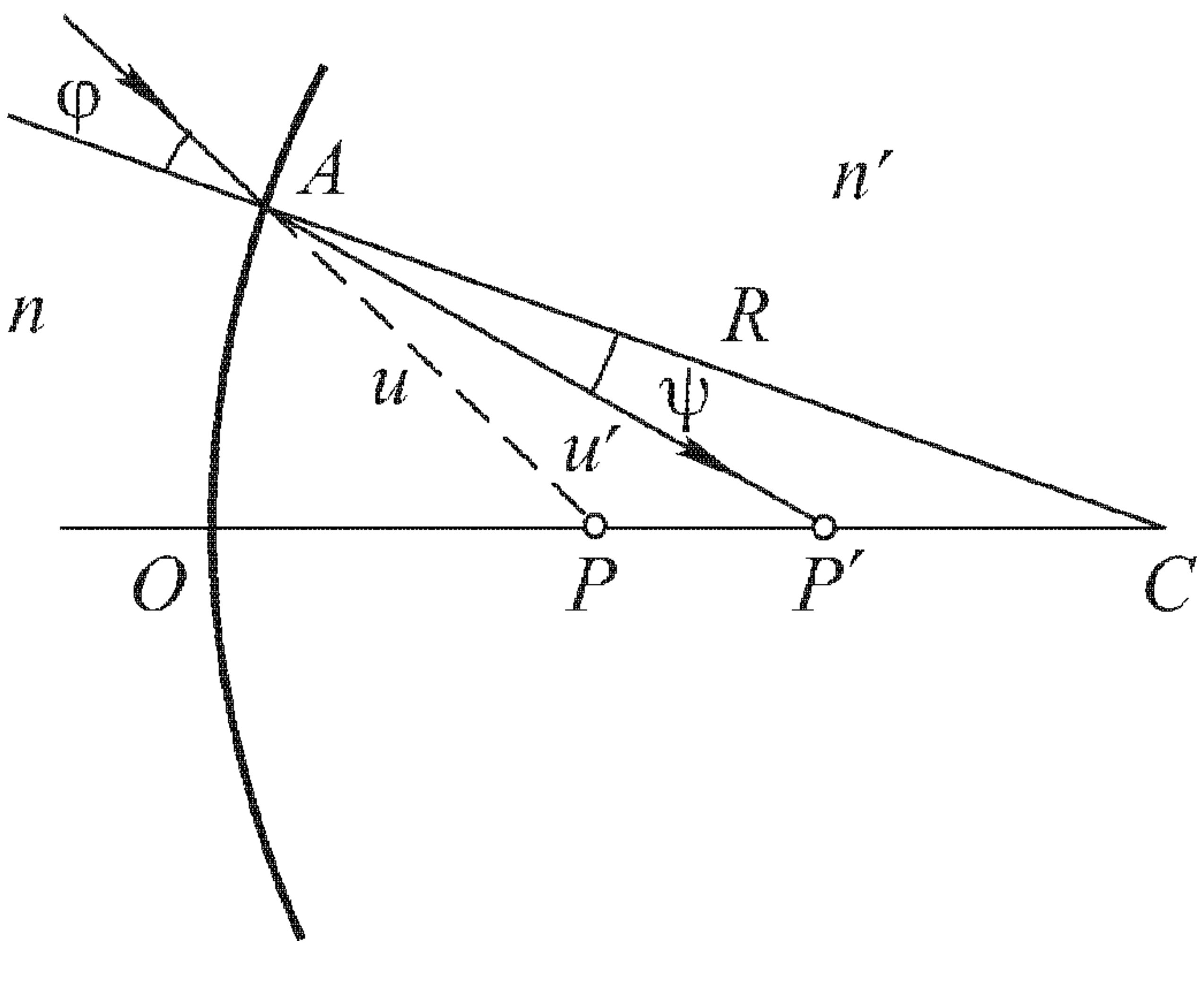


Рис. 40

Полагая формально $n' = -n$, из (10.2) получаем формулу для сферического зеркала:

$$\frac{1}{x} + \frac{1}{x'} = \frac{2}{R}. \tag{10.3}$$

3. Рассмотрим теперь случай, когда точечный источник P не лежит на главной оптической оси (рис. 41). Проведем прямую PC, соединяющую точку P с центром кривизны преломляющей поверхности. Такую прямую можно рассматривать как оптическую ось, сведя тем самым разбираемый случай к предыдущему. Для параксиальных лучей $O'P \approx OQ$, $O'P' \approx OQ'$

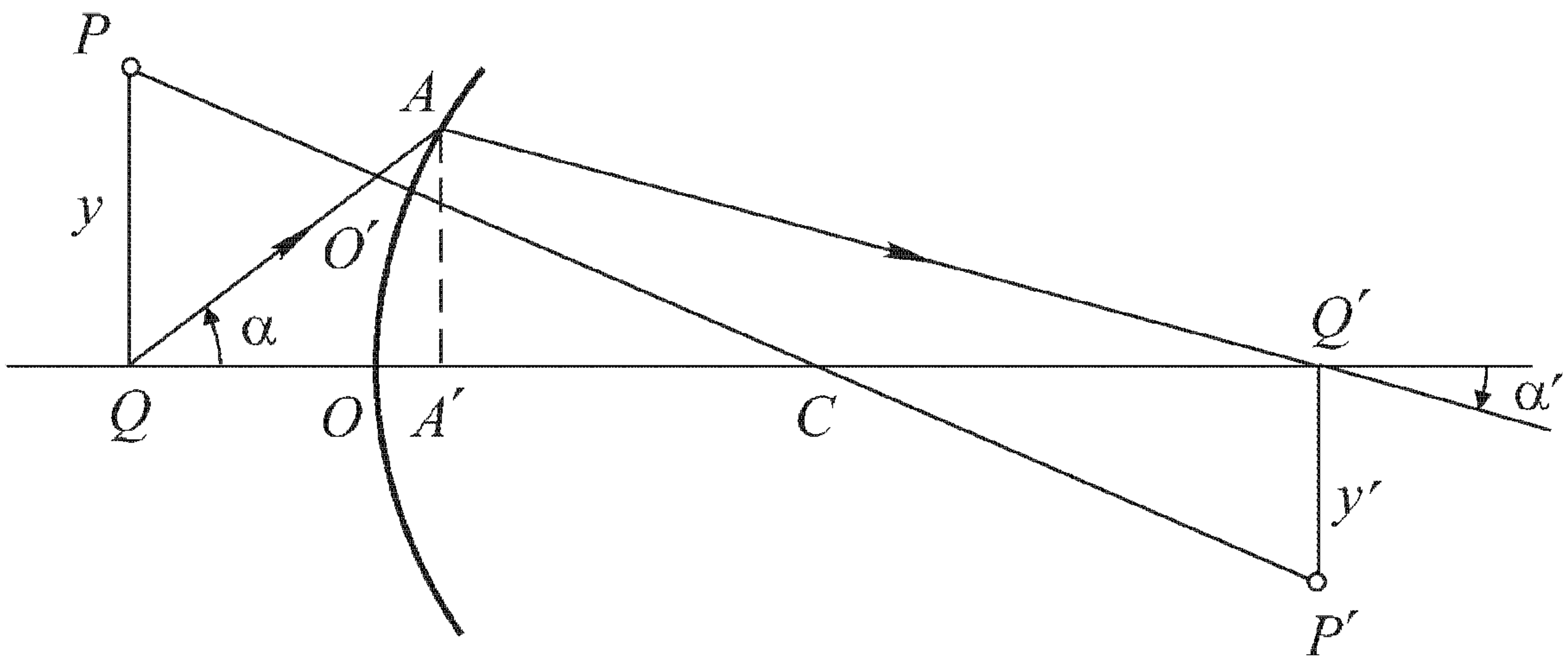


Рис. 41

(здесь Q и Q' — проекции точек P и P' на главную оптическую ось). Абсцисса x' точки P' определится из уравнения (10.2), а ордината y' — из соотношения

$$\frac{y'}{y} = -\left|\frac{CQ'}{CQ}\right| = \frac{x' - R}{x - R},$$

которое следует из подобия треугольников PCQ и $P'CQ'$. В результате находим

$$x' = \frac{n'Rx}{(n' - n)x + nR}, \quad y' = \frac{nRy}{(n' - n)x + nR}. \tag{10.4}$$

Проведем через точку Q произвольный параксиальный луч QA, образующий с главной оптической осью какой-то угол α. Угол наклона преломленного луча AQ' с той же осью обозначим через α'. Отсчет углов производится в направлении от главной оптической оси к лучу. Если это направление противоположно направлению вращения часовой стрелки, то угол считается положительным; в противоположном случае его следует считать отрицательным (на рис. 41 угол α положительный, угол α' отрицательный). Опустим из точки A перпендикуляр AA' на главную оптическую ось. В приближении параксиальной оптики его длину можно представить в виде $|AA'| = -x\alpha = -x'\alpha'$. Следовательно, $x\alpha = x'\alpha'$. Но из формул (10.4) следует

$$\frac{x'}{x} = \frac{n'y'}{ny}, \tag{10.5}$$

а потому

$$ny\alpha = n'y'\alpha'. \tag{10.6}$$

Таким образом, величина $ny\alpha$ не изменяется при преломлении параксиального луча на сферической поверхности. Эта величина называется *инвариантом Лагранжа–Гельмгольца*, а равенство (10.6) — *теоремой Лагранжа–Гельмгольца*. Теорема, очевидно, справедлива для центрированных систем, состоящих из какого угодно числа преломляющих и отражающих сферических поверхностей.

Формулы (10.4) можно положить в основу геометрической теории любых центрированных систем в параксиальных лучах. Применяя их к первой преломляющей поверхности сложной системы, найдем положение изображения, возникающее от преломления на этой поверхности. Полученное промежуточное изображение играет роль предмета для преломления на второй сферической поверхности. С помощью тех же формул (10.4) можно найти положение второго промежуточного изображения, возникающего от преломления на второй сферической поверхности, и т.д. В конце концов путем последовательного применения формул (10.4) к каждой из преломляющих поверхностей можно найти положение окончательного изображения, даваемого всей системой.

4. В качестве примера рассмотрим тонкую линзу, ограниченную сферическими поверхностями с радиусами кривизны R_1 и R_2. Относительный показатель преломления линзы обозначим через n, приняв за единицу показатель преломления окружающего пространства. Пусть точечный предмет P находится на главной оптической оси линзы. Толщиной линзы будем пренебрегать и поместим начало координат в ее центре. Обозначим через x абсциссу точки P, через x_1 — абсциссу ее промежуточного изображения P_1, возникающего от преломления лучей на первой поверхности линзы. Абсциссу x_1 можно найти с помощью формулы (10.2), если в ней сделать замену $n \to 1$, $n' \to n$, $x' \to x_1$, $R \to R_1$. Это дает

$$\frac{1}{x} - \frac{n}{x_1} = \frac{1-n}{R_1}.$$

Промежуточное изображение P_1 будем рассматривать как предмет при преломлении света на второй сферической поверхности линзы. Изображение точки P, возникающее при таком преломлении, и будет окончательным изображением P', которое дает линза. Абсцисса x' точки P' найдется из формулы (10.2), если в ней сделать замену $n' \to 1$, $x \to x_1$, $R \to R_2$. Таким путем находим

$$\frac{n}{x_1} - \frac{1}{x'} = \frac{n-1}{R_2}.$$

Складывая это равенство с предыдущим, получим

$$\frac{1}{x} - \frac{1}{x'} = -(n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \tag{10.7}$$

Это хорошо известная формула тонкой линзы

$$\frac{1}{x}-\frac{1}{x'}=-\frac{1}{f}. \tag{10.8}$$

Фокусное расстояние f определяется формулой

$$\frac{1}{f}=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right). \tag{10.9}$$

Эта формула справедлива для всяких тонких линз: двояковыпуклых, двояковогнутых, плоско-выпуклых и т.д. Надо только придерживаться правила знаков, сформулированного в п. 2 настоящего параграфа. Так, для двояковыпуклой линзы $R_1 > 0$, $R_2 < 0$, а потому фокусное расстояние f положительно. Для двояковогнутой линзы $R_1 < 0$, $R_2 > 0$, и фокусное расстояние f отрицательно. (В обоих случаях предполагается, что $n > 1$.)

§ 11. Общие свойства центрированных оптических систем

1. Свойства центрированных оптических систем в параксиальных лучах были систематически исследованы Гауссом (1777–1855) в 1841 г. Поэтому оптику параксиальных лучей часто называют *гауссовой оптикой*. При изложении относящихся сюда вопросов мы применим *аналитический метод*. Он менее нагляден, чем геометрический метод. Зато аналитический метод отличается большей простотой и систематичностью.

В случае одной преломляющей поверхности координаты x, y точки-предмета связаны с координатами x', y' точки-изображения формулами (10.4). В этих формулах используется одна и та же координатная система в пространствах предметов и изображений. Выберем теперь в этих пространствах *разные системы координат*, получающиеся из исходной системы параллельным переносом вдоль главной оптической оси. Начала координат этих систем лежат на главной оптической оси, но могут и не совпадать друг с другом. При новом выборе координатных систем формулы (10.4) преобразуются в

$$x'=\frac{ax+b}{cx+d},\qquad y'=\frac{e}{cx+d}y,$$

где a, b, c, d, e — постоянные. Их легко вычислить, если известны положения начал координат и параметры преломляющей сферической поверхности. Этими формулами устанавливается соответствие между точками пространства предметов и пространства изображений. Оно называется *коллинеарным со-*

ответствием (точнее — *коллинеарным соответствием с осевой симметрией*).

Допустим теперь, что после прохождения через первую преломляющую поверхность лучи испытывают преломление на второй сферической поверхности. Тогда получится второе изображение — точка P'' с координатами x'', y''. Они связаны с x', y' формулами такого же вида, т.е.

$$x'' = \frac{a'x' + b'}{c'x' + d'}, \quad y'' = \frac{e'}{c'x' + d'}\,y'$$

с новыми коэффициентами a', b', c', d', e'. Исключим из этих и предыдущих соотношений координаты x', y' промежуточного изображения P'. Нетрудно убедиться, что таким путем снова получатся формулы коллинеарного соответствия:

$$x'' = \frac{a''x + b''}{c''x + d''}, \quad y'' = \frac{e''}{c''x + d''}\,y,$$

причем коэффициенты a'', b'', c'', d'', e'' могут быть выражены через десять коэффициентов a, b, ..., d', e'. Таким образом, *два или несколько последовательно выполненных коллинеарных соответствий эквивалентны одному коллинеарному соответствию*. Следовательно, *любая центрированная система в параксиальных лучах устанавливает коллинеарное соответствие между точками пространства предметов и пространства изображений*. Его можно выразить формулами

$$x' = \frac{ax + b}{cx + d}, \quad y' = \frac{e}{cx + d}\,y, \quad z' = \frac{e}{cx + d}\,z. \tag{11.1}$$

Здесь добавлена формула для координат z и z'. Ввиду осевой симметрии, она совпадает с соответствующей формулой для y и y'.

Коллинеарное соответствие определяется *четырьмя параметрами*, за которые можно принять отношения четырех из коэффициентов a, b, c, d, e к пятому. Поэтому и произвольная центрированная система характеризуется также *четырьмя параметрами*.

Разрешая уравнения (11.1) относительно x, y, z, получим

$$x = \frac{a'x' + b'}{c'x' + d'}, \quad y = \frac{e'}{c'x' + d'}\,y', \quad z = \frac{e'}{c'x' + d'}\,z', \tag{11.2}$$

причем

$$b' = b, \quad c' = c, \quad a' = -d, \quad d' = -a, \quad e' = \frac{bc - ad}{e}. \tag{11.3}$$

Обратное преобразование, таким образом, выражается также формулами коллинеарного соответствия, что, очевидно, является следствием обратимости светового пути (см. § 10, п. 1).

2. Из формул коллинеарного соответствия вытекают следующие свойства оптических изображений в центрированных системах.

1. *Каждая плоскость пространства предметов изображается в виде плоскости.* Действительно, уравнение плоскости в пространстве предметов

$$Ax + By + Cz + D = 0$$

после замены координат x, y, z их выражениями через x', y', z' по формулам (11.2) переходит в уравнение вида

$$A'x' + B'y' + C'z' + D' = 0,$$

которое представляет плоскость в пространстве изображений.

2. *Каждая прямая пространства предметов изображается в виде прямой*, так как прямую можно рассматривать как линию пересечения двух плоскостей.

3. *Каждая точка пространства предметов изображается в виде точки*, так как всякую точку можно рассматривать как точку пересечения трех плоскостей.

3. Из формул (11.1) следует, что конечным значениям x, y, z соответствуют, вообще говоря, конечные значения x', y', z'. Исключение составляют точки плоскости

$$cx + d = 0. \tag{11.4}$$

Каждая точка такой плоскости изображается бесконечно удаленной точкой. Это означает, что все лучи, вышедшие из одной и той же точки плоскости (11.4), после прохождения через оптическую систему становятся параллельными. Плоскость (11.4) называется *фокальной плоскостью пространства предметов, или передней фокальной плоскостью оптической системы.* Аналогично, плоскость

$$c'x' + d' = 0 \tag{11.5}$$

называется *фокальной плоскостью пространства изображений*, или *задней фокальной плоскостью оптической системы.* Параллельные лучи после прохождения через оптическую систему пересекаются в одной из точек этой плоскости.

Точки пересечения фокальных плоскостей с главной оптической осью называются *фокальными точками*, или *главными фокусами оптической системы.* Главный фокус пространства предметов (*передний главный фокус*) будем обозначать через F, а главный фокус пространства изображений (*задний главный фокус*) — через F'. Согласно формулам (11.4), (11.5) и (11.3), абсциссы главных фокусов системы определяются выражениями:

$$x_F = -\frac{d}{c}, \quad x'_{F'} = -\frac{d'}{c'} = \frac{a}{c}. \tag{11.6}$$

Система может и не иметь фокальных плоскостей. Это будет, когда $c = 0$, и следовательно, $c' = 0$. Такие системы называются *афокальными*, или *телескопическими*. Они являются предельными случаями обычных систем, когда обе фокальные плоскости сдвинуты в бесконечность. После прохождения через афокальную систему всякий параллельный пучок лучей остается параллельным, могут изменяться лишь ширина и направление пучка. Примером афокальной системы может служить зрительная труба (телескоп), установленная на бесконечность. В этом случае задняя фокальная плоскость объектива совмещается с передней фокальной плоскостью окуляра.

Рассмотрим сначала системы с конечными фокусными расстояниями.

4. Отношение y'/y называется *поперечным увеличением*, или просто *увеличением системы*. Согласно формулам (11.1) или (11.2), оно не зависит от y и z. Отсюда следует, что *изображение плоского предмета, перпендикулярного к главной оптической оси, подобно самому предмету*. Если увеличение положительное, то изображение *прямое*. В противоположном случае изображение *обратное*.

Две сопряженные плоскости, отображающиеся друг в друга с поперечным увеличением $y'/y = +1$, называются *главными плоскостями оптической системы*. Уравнения таких плоскостей можно получить, полагая в формулах (11.1) и (11.2) $y' = y$. Это дает

$$cx + d - e = 0, \quad c'x' + d' - e' = 0. \tag{11.7}$$

Первая плоскость называется *главной плоскостью пространства предметов* (*передней главной плоскостью*), вторая — *главной плоскостью пространства изображений* (*задней главной плоскостью*). Точки пересечения главных плоскостей с главной оптической осью называются *главными точками центрированной системы*. Главную точку в пространстве предметов будем обозначать через H, а в пространстве изображений — через H'. Полагая в формулах (11.1) и (11.2) $y = y'$, находим абсциссы главных точек:

$$x_H = \frac{e-d}{c}, \quad x'_{H'} = \frac{e'-d'}{c'} = \frac{bc + a(e-d)}{ec}. \tag{11.8}$$

5. Главные и фокальные точки центрированной системы называются ее *кардинальными точками*. Так как коллинеарное соответствие определяется четырьмя параметрами, то положение четырех кардинальных точек полностью определяет коллинеарное соответствие. Фокальные и главные точки полностью характеризуют оптическую систему в том смысле, что, зная положение этих точек, можно найти изображение любого предмета, даваемое оптической системой.

Допустим сначала, что предмет P точечный и не лежит на главной оптической оси системы (рис. 42). Опустим из точки P

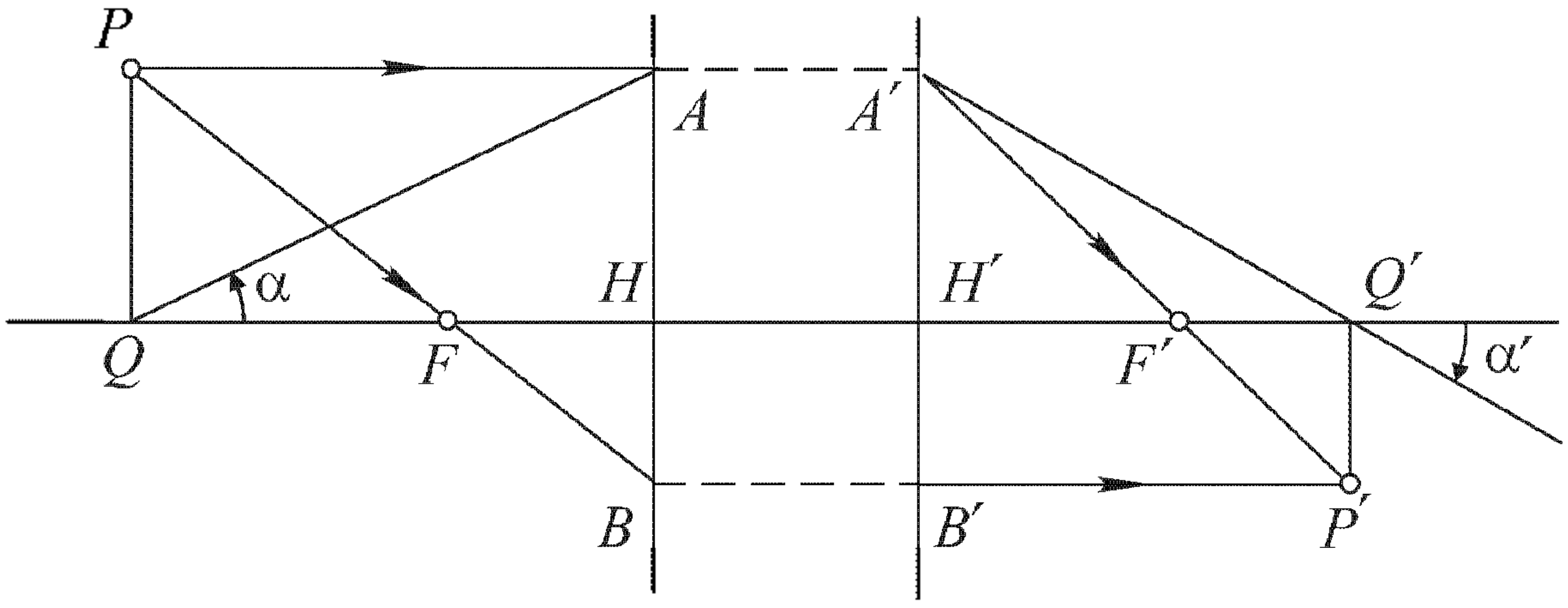


Рис. 42

перпендикуляр PQ на эту ось. Луч PA, параллельный главной оптической оси, или его продолжение в сторону пространства изображений встретит главные плоскости в сопряженных точках A и A'. Поэтому после прохождения через оптическую систему этот луч или его продолжение пройдет через точку A'. Кроме того, он должен пройти через задний фокус F'. Двумя точками A' и F' положение луча в пространстве изображений определяется полностью. Проведем теперь второй луч PF, проходящий через передний фокус F. Он или его продолжение встретит переднюю главную плоскость в точке B. Проведя через B прямую $BB'P'$, параллельную главной оптической оси, найдем положение луча $B'P'$ в пространстве изображений. Точка P', в которой пересекаются лучи $A'P'$ и $B'P'$, и будет изображением точки P. При этом проекция Q точки P на главную оптическую ось изобразится проекцией точки P' на ту же ось. Это дает правило построения изображения точки и в том случае, когда она лежит на главной оптической оси. Отрезок QP изображается отрезком $Q'P'$.

Расстояния главных точек от соответствующих фокальных точек называются *главными фокусными расстояниями* системы. Фокусное расстояние в пространстве предметов будем обозначать через f, в пространстве изображений — через f'. В соответствии с принятым нами правилом знаков фокусное расстояние считается положительным, если падающий свет идет в направлении от главного фокуса к соответствующей главной плоскости. Фокусные расстояния определяются формулами

Надлежащим выбором начал координат можно упростить формулы (11.1) и (11.2). Рассмотрим два важнейших случая.

6. Случай 1. Начала координат помещены в главные точки H и H'. Абсциссы в этих координатных системах будем обозначать через ξ и ξ', а ординаты — через η и η'. Очевидно, $\eta = y$, $\eta' = y'$, а по самому выбору начал координат $\xi_H = \xi'_{H'} = 0$. Поэтому из (11.8) следует: $e = d$, $b = 0$. Формулы (11.1) преобразуются в

$$\xi' = \frac{a\xi}{c\xi + d}, \quad \eta' = \frac{d}{c\xi + d}\,\eta.$$

Абсциссы фокальных точек, ввиду соотношений (11.6), будут $\xi_F = -d/c$, $\xi'_{F'} = a/c$, а фокусные расстояния $f = d/c$, $f' = -a/c$. Исключая с помощью этих соотношений коэффициенты a, c, d, получим

$$\xi' = \frac{f'\xi}{f - \xi}, \quad \eta' = \frac{f\eta}{f - \xi}. \tag{11.10}$$

Первую из этих формул можно записать также в виде

$$\frac{f}{\xi} + \frac{f'}{\xi'} = -1. \tag{11.11}$$

Как видно из рис. 42, $HA = -\xi\alpha$, $H'A' = -\xi'\alpha'$. Так как $HA = H'A'$, то $\xi\alpha = \xi'\alpha'$. Далее, из уравнений (11.10) находим: $f\eta\xi' = -f'\eta'\xi$, так что

$$f\eta\alpha = -f'\eta'\alpha'. \tag{11.12}$$

Но по теореме Лагранжа–Гельмгольца $n\eta\alpha = n'\eta'\alpha'$, а потому

$$\frac{f}{n} = -\frac{f'}{n'}. \tag{11.13}$$

Такова связь между фокусными расстояниями f и f'.

Допустим, как это обычно бывает, что среда по обе стороны оптической системы одна и та же. Если система содержит четное число отражений (такая система называется *диоптрической*), то $n = n'$, и следовательно, $f = -f'$. В этом случае формула (11.11) переходит в *формулу линзы*:

$$\frac{1}{\xi} - \frac{1}{\xi'} = -\frac{1}{f}. \tag{11.14}$$

Если же система содержит нечетное число отражающих поверхностей (такая система называется *катоптрической*), то $n = -n'$, и следовательно, $f = f'$. Тогда формула (11.11) переходит в *формулу зеркала*:

$$\frac{1}{\xi} + \frac{1}{\xi'} = -\frac{1}{f}. \tag{11.15}$$

В диоптрических системах направления распространения падающего и прошедшего света одинаковы, а в катоптрических противоположны.

7. Случай 2. Начала координатных систем помещены в главные фокусы F и F'. Абсциссы в этих системах будем обозначать через X и X', а ординаты через Y и Y'. Очевидно, $Y = y = \eta$, $Y' = y' = \eta'$. Кроме того, по самому выбору начал координат $X_F = X'_{F'} = 0$. Следовательно, должно быть $a = d = 0$, $b/c = ff'$, $e/c = f$, как в этом нетрудно убедиться с помощью формул (11.6) и (11.9). В результате формулы (11.1) переходят в

$$XX' = ff', \tag{11.16}$$

$$\frac{Y'}{Y} = \frac{f}{X} = \frac{X'}{f'}. \tag{11.17}$$

Эти простые формулы были известны еще Ньютону, — правда, для частных случаев.

8. К кардинальным точкам, наряду с главными и фокальными, относят еще *узловые точки*. Они естественно появляются при рассмотрении *углового увеличения* системы. Угловое увеличение есть отношение α'/α углов α' и α, образуемых с главной оптической осью проходящим и падающим лучами. *Узловыми точками называются две сопряженные точки K и K', лежащие на главной оптической оси*, которые отображаются друг в друга с угловым увеличением $+1$. Это значит, что всякий луч, проходящий через узловую точку K, после прохождения через оптическую систему остается параллельным своему исходному направлению и проходит через вторую узловую точку системы K'. Для доказательства существования и определения положения узловых точек положим в формуле (10.6) $\alpha = \alpha'$. Это дает $ny = n'y'$, или на основании формулы (11.13) $y'/y = -f/f'$. Сравнивая этот результат с формулами (11.17), находим абсциссы узловых точек относительно фокальных точек:

$$X_K = -f', \quad X'_{K'} = -f, \tag{11.18}$$

или на основании определения фокусных расстояний

$$FK = H'F', \quad F'K' = HF. \tag{11.19}$$

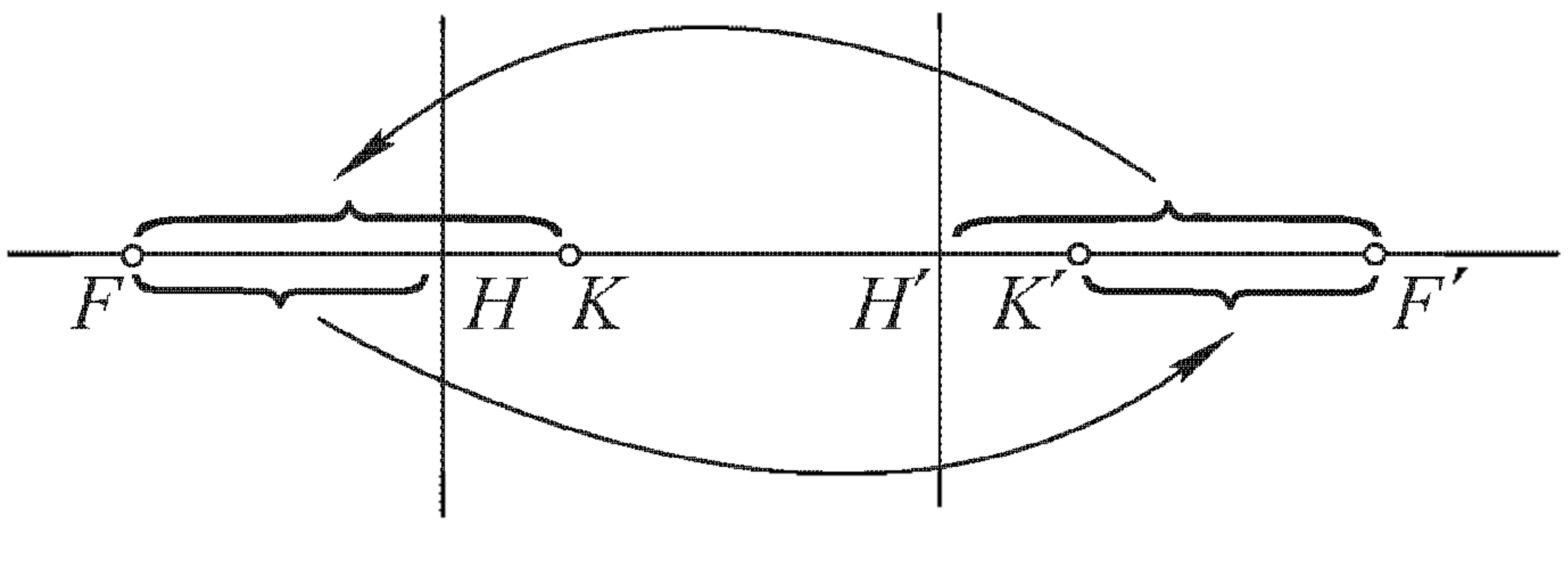


Рис. 43

Это приводит к следующему правилу для нахождения узловых точек. Отрезки HF и $H'F'$ надо перенести параллельно самим себе, чтобы их начальные точки H и H' совпали с фокусами F и F' соответственно. Тогда другие концы этих отрезков F и F' укажут положение узловых точек K' и K (рис. 43). Если $n' = n$, то $f' = -f$.

В этом случае узловые точки совпадают с главными. Узловые точки также могут быть использованы для построения оптических изображений в центрированных системах.

Иногда к числу кардинальных точек относят *обратные главные* и *обратные узловые точки*. Первые характеризуются линейным увеличением, равным минус единице; вторые — угловым увеличением, равным также минус единице. Доказательство существования и определение положения таких пар точек не представляют затруднений.

9. Отношение длины $\delta X'$ изображения бесконечно малого отрезка, параллельного главной оптической оси, к длине δX самого отрезка, называется *осевым* или *продольным увеличением*. Для этого увеличения из формулы (11.16) находим

$$\frac{\delta X'}{\delta X} = -\frac{X'}{X} = -\frac{ff'}{X^2} = -\frac{X'^2}{ff'}. \tag{11.20}$$

Сравнение этих формул с формулами (11.17) показывает, что *осевое увеличение в общем случае не равно поперечному увеличению*. Отсюда следует, что *изображение бесконечно малого объемного предмета в центрированной оптической системе, вообще говоря, не подобно самому предмету. Исключение составляет случай, когда предмет помещен в одной из узловых или в одной из обратных узловых точек системы.* В самом деле, для изображения бесконечно малого объемного предмета с сохранением подобия необходимо и достаточно, чтобы осевое увеличение по абсолютной величине было равно поперечному увеличению, т.е.

$$\frac{ff'}{X^2} = \pm\frac{f}{X}.$$

Отсюда $X = \pm f'$. Формулы (11.17) и (11.13) дают

$$\frac{Y'}{Y} = \pm\frac{f}{f'} = \mp\frac{n}{n'}.$$

После этого из теоремы Лагранжа–Гельмгольца (10.6) получаем $\alpha = \pm\alpha'$, что и требовалось доказать.

Если система диоптрическая (см. п. 6), то знаки f и f' противоположны. В этом случае, как видно из (11.17), δX и $\delta X'$ имеют одинаковые знаки. Отсюда следует, что при перемещении предмета вдоль оптической оси его изображение перемещается *в том же направлении*. Напротив, если система катоптрическая, то при перемещении предмета вдоль оптической оси его изображение перемещается *в противоположном направлении*. А так как в диоптрических системах направления распространения света в пространствах предметов и изображений одинаковы, а в катоптрических противоположны, то для всех систем справедливо такое правило:

Если предмет перемещается вдоль оптической оси в направлении распространения падающего света, то его изображение перемещается в направлении распространения прошедшего света, и наоборот.

10. В случае телескопической системы $c = 0$. Но коэффициент d в нуль обращаться не должен, иначе формулы (11.1) потеряли бы смысл. Не должны обращаться в нуль также коэффициенты a и e, так как в противном случае из формул (11.1) мы получили бы $x' = \mathrm{const}$, $y' = 0$, $z' = 0$, т.е. любая точка изображалась бы всегда одной и той же точкой. Такой случай в оптических системах не встречается, он имеет только формально-математический характер. Итак, для телескопической системы $c = 0$, $d \neq 0$, $a \neq 0$, $e \neq 0$. Следовательно, должно быть также $c' = 0$, $d' \neq 0$, $a' \neq 0$, $e' \neq 0$, как это видно из формул (11.3). Телескопическую систему можно рассматривать как предельный случай фокальной системы, обе фокальные точки которой удалены в бесконечность ($x_F = \pm\infty$, $x'_{F'} = \pm\infty$). Оба фокусных расстояния f и f' телескопической системы бесконечно велики, хотя их отношение $f/f' = -n/n'$ и остается конечным.

Формулы (11.1) для телескопической системы принимают вид

$$x' = Ax + C, \quad y' = By, \tag{11.21}$$

где $A = a/d$, $B = e/d$, $C = b/d$ — постоянные. Отсюда видно, что *всякий параллельный пучок света после прохождения через телескопическую систему остается параллельным.* Постоянные A и B не независимы. Действительно, формулы (11.9) для телескопических систем дают

$$\frac{f}{f'} = -\frac{e^2}{ad} = -\frac{B^2}{A},$$

или на основании (11.13)

$$\frac{B^2}{A} = \frac{n}{n'}. \tag{11.22}$$

Если за начала координатных систем принять любую пару сопряженных точек оптической оси, то формулы (11.21) упростятся и перейдут в

$$x' = Ax, \quad y' = By. \tag{11.23}$$

Таким образом, *поперечное и осевое увеличения телескопической системы постоянны, т.е. не зависят от положения предмета.* Постоянно также и угловое увеличение, так как по формуле (10.6)

$$\frac{\alpha'}{\alpha} = \frac{ny}{n'y'} = \frac{n}{Bn'} = \frac{B}{A}. \tag{11.24}$$

Для зрительных труб и любых систем, у которых $n = n'$,

$$\frac{\alpha'}{\alpha} = \frac{y}{y'}. \tag{11.25}$$

В этом случае угловое увеличение называют просто *увеличением* трубы, опуская прилагательное «угловое». Эта величина показывает, во сколько раз угол, под которым виден бесконечно удаленный малый предмет в трубу, больше угла, под которым он был бы виден невооруженным глазом. Согласно (11.25), угловое увеличение зрительной трубы равно обратному значению ее поперечного увеличения. Отсюда следует, что *увеличение зрительной трубы численно равно отношению ширины падающего пучка лучей к ширине соответствующего выходящего пучка.*

Поясним последнее утверждение на примере *кеплеровой трубы*, состоящей из двух тонких собирающих линз — объектива O и окуляра O' (рис. 44). Задняя фокальная плоскость объектива совпадает с передней фокальной плоскостью окуляра. Пусть в трубу рассматривается бесконечно удаленный предмет. Объектив дает изображение FC этого предмета в общей фокальной плоскости объектива и окуляра. Допустим для простоты, что один из крайних лучей (луч AF) идет вдоль оптической оси трубы, а другой (луч BC) — под углом α к ней. Проходя через оптический центр объектива, луч BC не испытывает преломления. Угол α есть угол, под которым бесконечно удаленный предмет виден невооруженным глазом; угол α' — угол, под которым тот же предмет виден в зрительную трубу. *Угловое увеличение* α'/α, согласно рис. 44, *численно равно отношению фокусного расстояния объектива к фокусному расстоянию окуляра.* Но это отношение в свою очередь равно отношению ширины падающего параллельного пучка лучей к ширине соответствующего выходящего пучка (рис. 45).

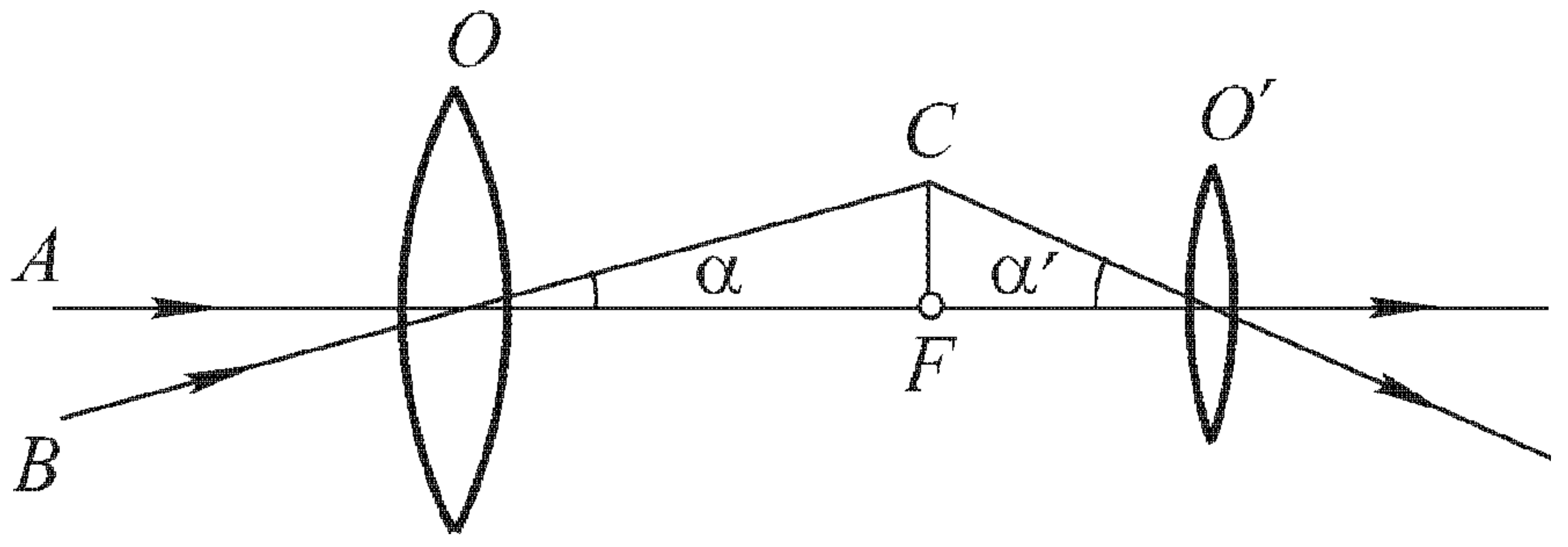


Рис. 44

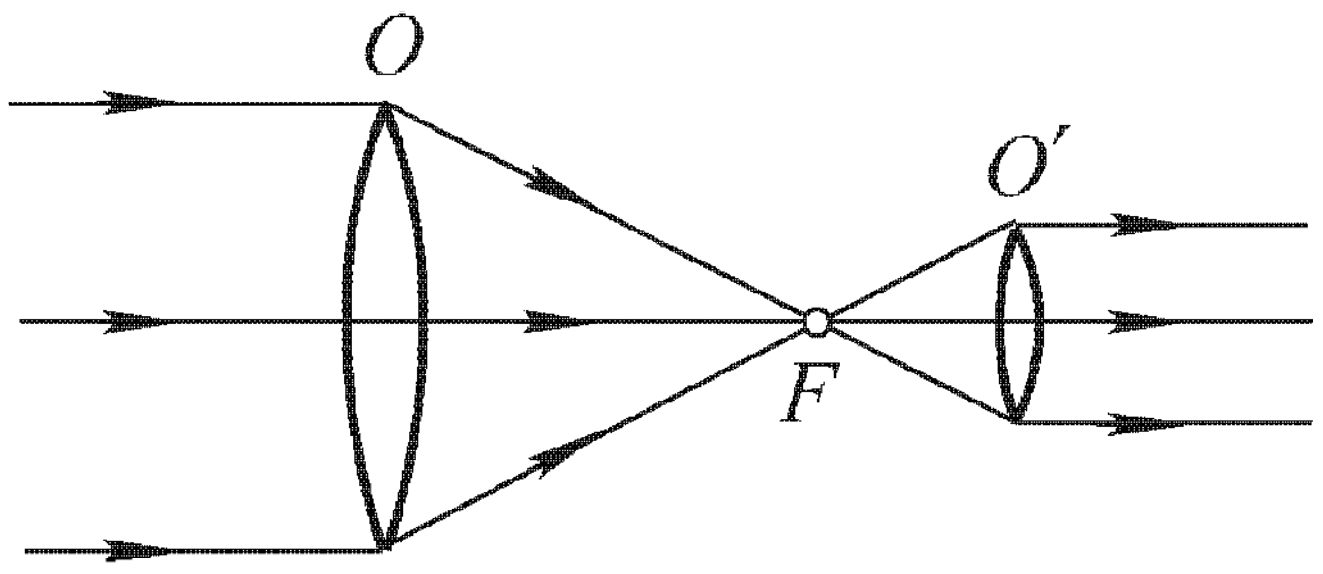


Рис. 45

11. Более общий подход к вопросу об угловом увеличении телескопической системы дает *принцип Ферма.* Этот подход применим и к телескопическим системам, не обладающим осевой симметрией. *Телескопической системой в общем случае называют любую оптическую систему, при прохождении через которую каждый параллельный пучок света остается параллель-*

ным. Пусть AB (рис. 46) — плоский участок волнового фронта перед телескопической системой. После прохождения через эту систему (не изображенную на рисунке) он переходит в плоский участок $A'B'$. Продолжением луча AC является луч $C'A'$, а луча BD — луч $D'B'$. Таким образом, $(ACC'A') = (BDD'B')$. Возьмем другой плоский участок волнового фронта AE, наклоненный к AB под бесконечно малым углом α. За оптической системой волновой фронт AE перейдет в волновой фронт $A''E'$, образующий с прежним волновым фронтом $A'B'$ угол α'. Угловое увеличение системы будет $N = \alpha'/\alpha$. Возьмем тот волновой фронт $A''E'$, который проходит через точку A'. На основании следствия принципа Ферма, доказанного в п. 4 § 7, оптические длины лучей AA'' и EE' при смещении вдоль волнового фронта $A''E'$ будут меняться во втором или высшем порядке малости. В этом порядке указанные оптические длины останутся неизменными, если заменить их оптическими длинами $(ACC'A')$ и $(EDD'B'K)$. Таким образом, $(ACC'A') = (EDD'B'K)$. Сравнивая это соотношение с предыдущим, получим $(BEDD'B') = (EDD'B'K)$, откуда $(BE) = (B'K)$, или

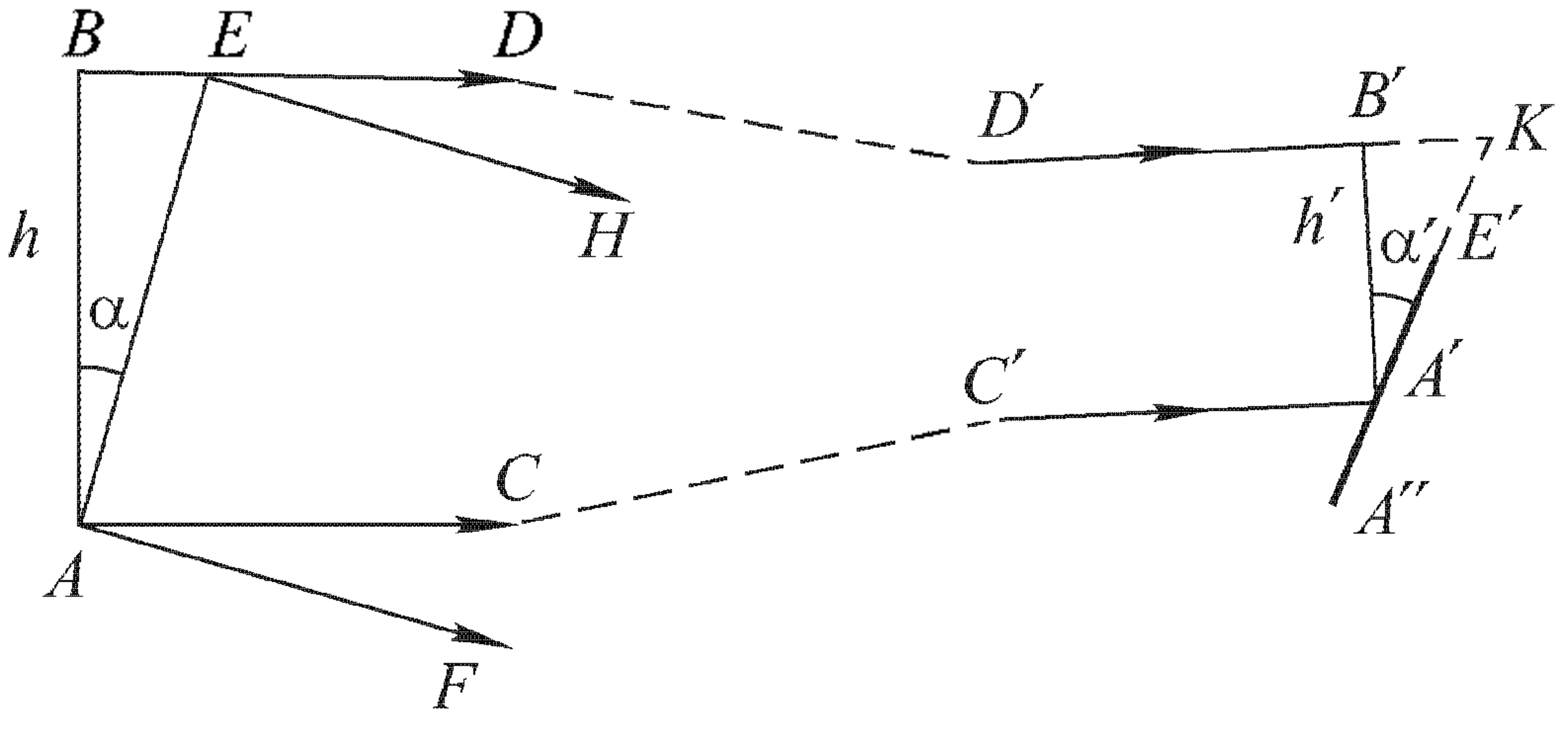


Рис. 46

$$nh\alpha = n'h'\alpha', \tag{11.26}$$

где n — показатель преломления среды перед телескопической системой, n' — за этой системой, h — поперечное сечение падающего пучка лучей, а h' — выходящего. Если, как обычно бывает, $n' = n$, то

$$N = \frac{\alpha'}{\alpha} = \frac{h}{h'}, \tag{11.27}$$

т.е. *увеличение телескопической системы равно отношению ширин пучков света до и после прохождения через эту систему*. Полученные результаты имеют общий характер и не зависят от конкретного устройства телескопической системы.

Увеличение может и не быть одинаковым по всем направлениям. Рассмотрим, например, призму, преломляющее ребро которой вертикально. Вертикальные размеры параллельного светового пучка после прохождения через призму не изменяются, тогда как горизонтальные размеры, вообще говоря, меняются. Призма будет давать увеличение в горизонтальном направлении, но не будет увеличивать в вертикальном направлении. Изображение, даваемое призмой, получится вытянутым или сплюснутым в горизонтальном направлении, в зависимости от того, уменьшаются или увеличиваются поперечные размеры светового пучка в этом направлении. Искажения не получится, если держать призму под углом наименьшего отклонения, но в этом случае не будет и увеличения. Комбинацией двух призм, преломляющие ребра которых взаимно перпендикулярны, можно произвольно увеличивать или уменьшать поперечное сечение светового пучка и притом одинаково по всем направлениям. Такая комбинация действует как обыкновенная зрительная труба. Вращая каждую из призм вокруг своего преломляющего ребра, можно получить произвольное увеличение. Недостатками такого оптического прибора являются большие геометрические и хроматические аберрации, вносимые им. Практического значения он не имеет.

§ 12. Сложение центрированных систем. Толстые линзы

1. Пусть две центрированные системы соединены вместе таким образом, что их оптические оси совпадают. Если известны параметры каждой системы, а также их взаимное расположение, то геометрическим построением или аналитическим расчетом можно определить положение всех кардинальных точек сложной оптической системы, состоящей из этих двух систем.

Обозначим фокусные расстояния первой системы через f_1 и f_1', а второй системы — через f_2 и f_2'. Пусть Δ означает расстояние передней фокальной точки F_2 второй системы от задней фокальной точки F_1' первой системы (рис. 47). Это расстояние

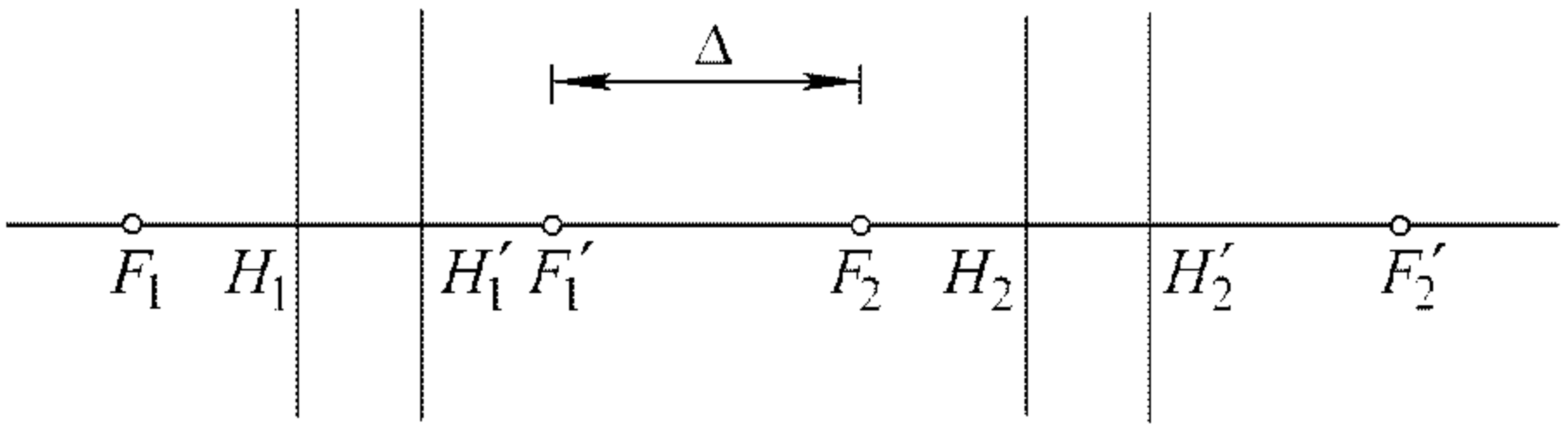


Рис. 47

называется *оптическим интервалом* двух систем и в соответствии с принятым правилом знаков считается положительным, если падающий свет идет в направлении от фокуса F_1' к фокусу F_2, в противоположном случае оптический интервал считается отрицательным. Заданием оптического интервала полностью определяется взаимное расположение складываемых систем. С целью упрощения вычислений начала координат для каждой из складываемых систем поместим в ее фокальные точки. Фокус F_1 примем за начало координат в пространстре предметов всей сложной системы, а фокус F_2' — за начало координат в пространстве изображений той же системы. Пусть x, y — координаты предмета, а x_1, y_1 — его изображения, даваемого первой из складываемых систем. Тогда

$$xx_1 = f_1 f_1', \qquad \frac{y_1}{y} = \frac{f_1}{x}.$$

Примем это промежуточное изображение за «предмет» для второй из складываемых систем. Координаты этого предмета в координатной системе с началом в точке F_2 будут $x_2 = x_1 - \Delta$, $y_2 = y_1$. Если x', y' — координаты изображения, даваемого второй системой (а следовательно, и всей сложной системой) относительно начала F_2', то

$$x_2 x' = f_2 f_2', \qquad \frac{y_1}{y_2} = \frac{x'}{f_2'}.$$

Исключая промежуточные координаты x_1, y_1, x_2, y_2, получим

$$x' = \frac{f_2 f_2'}{f_1 f_1' - \Delta \cdot x}\, x, \qquad y' = \frac{f_1 f_2}{f_1 f_1' - \Delta \cdot x}\, y. \tag{12.1}$$

Это — формулы коллинеарного соответствия с коэффициентами

$$a = f_2 f_2', \qquad b = 0, \qquad c = -\Delta, \qquad d = f_1 f_1', \qquad e = f_1 f_2.$$

Из них по формулам (11.6) и (11.9) находим координаты фокальных точек и фокусные расстояния сложной системы:

$$x_F = \frac{f_1 f_1'}{\Delta}, \qquad x'_{F'} = -\frac{f_2 f_2'}{\Delta}, \tag{12.2}$$

$$f = -\frac{f_1 f_2}{\Delta}, \qquad f' = \frac{f_1' f_2'}{\Delta}. \tag{12.3}$$

Координаты главных точек определяются выражениями

$$\begin{aligned} x_H &= x_F + f = f_1 \frac{f_1' - f_2}{\Delta}, \\ x'_{H'} &= x'_{F'} + f' = f_2' \frac{f_1' - f_2}{\Delta}. \end{aligned} \tag{12.4}$$

Следовательно,

$$\frac{x_H}{x'_{H'}} = \frac{f_1}{f_2'}. \tag{12.5}$$

Если оптический интервал Δ обращается в нуль, то фокусные расстояния f и f' обращаются в бесконечность, т.е. система будет *телескопической*. (Такой случай осуществляется, например, в зрительной трубе.) В этом случае уравнения (12.1) переходят в (11.23), причем

$$A = \frac{f_2 f_2'}{f_1 f_1'}, \quad B = \frac{f_2}{f_1'}.$$

Угловое увеличение сложной системы будет

$$\frac{\alpha'}{\alpha} = \frac{B}{A} = \frac{f_1}{f_2'}. \tag{12.6}$$

В частности, при $n = n'$

$$\frac{\alpha'}{\alpha} = -\frac{f_1}{f_2}. \tag{12.7}$$

Это согласуется с результатом, полученным выше для кеплеровой трубы.

Если обе складываемые системы телескопические, то составная система также телескопическая, причем ее угловое увеличение равно произведению угловых увеличений складываемых систем. Наконец, *соединение телескопической системы и системы с конечными фокусными расстояниями образует систему с конечными фокусными расстояниями при всякой последовательности расположения обеих систем.*

2. Величина, обратная главному фокусному расстоянию f' пространства изображений, взятая с противоположным знаком, т.е. $-1/f'$, называется *оптической силой системы*. Оптическая сила измеряется *диоптриями*. Диоптрия есть оптическая сила такой системы, фокусное расстояние $|f'|$ которой равно одному метру. Для собирательных тонких линз оптическая сила *положительна*, для рассеивающих *отрицательна*.

Определим оптическую силу сложной системы, зная оптические силы составляющих систем и их взаимное расположение. Будем предполагать, что показатели преломления всех пространств предметов и изображений одинаковы. Обозначим через l_{12} расстояние $H_1'H_2$ передней главной плоскости H_2 второй системы от задней главной плоскости H_1' первой системы. Оптический интервал между рассматриваемыми системами будет

$$\Delta = F_1'F_2 = F_1'H_1' + H_1'H_2 + H_2F_2 = f_1' + l_{12} - f_2 = f_1' + l_{12} + f_2'.$$

Подставляя это значение в формулу (12.3), получим

$$\frac{1}{f'} = \frac{1}{f_1'} + \frac{1}{f_2'} + \frac{l_{12}}{f_1' f_2'}. \tag{12.8}$$

В частности, когда задняя главная плоскость первой системы совпадает с передней главной плоскостью второй системы, то

$$\frac{1}{f'} = \frac{1}{f_1'} + \frac{1}{f_2'}, \tag{12.9}$$

т.е. *оптическая сила сложной системы равна сумме оптических сил составляющих систем.* Это имеет место, например, для двух тонких линз, прижатых вплотную одна к другой.

Применим полученные результаты к системе двух центрированных тонких линз: собирательной и рассеивающей, поставленных друг за другом. Пусть фокусные расстояния линз по абсолютной величине одинаковы: $f_1 = -f_2$, а потому $f_1' = -f_2'$. Оптический интервал между линзами $\Delta = f_1' + l_{12} + f_2' = l_{12}$, т.е. положителен ($l_{12} > 0$). Из формулы (12.6) получаем

$$f = -f' = -\frac{f_1' f_2'}{l_{12}} = -\frac{f_1 f_2}{l_{12}} = \frac{f_1^2}{l_{12}} > 0.$$

Далее, по формулам (12.2) и (12.4) находим $x_H = x'_{H'} = 0$, т.е. главные плоскости H и H' системы проходят соответственно через фокусы F_1 и F_1' и находятся на расстоянии l_{12} друг от друга. Абсцисса главного фокуса F' системы:

$$x'_{F'} = -\frac{f_2 f_2'}{l_{12}} = \frac{f_1^2}{l_{12}},$$

абсцисса линзы L_2:

$$x'_{L_2} = f_2' = -f_1' = f_1.$$

Таким образом,

$$x'_{F'} - x'_{L_2} = \frac{f_1^2}{l_{12}} - f_1.$$

Для того чтобы система линз собирала лучи, параллельные главной оптической оси, в действительном фокусе, т.е. была собирающей, необходимо, чтобы эта разность была положительна. Если первая линза рассеивающая, то указанное условие соблюдается всегда, так как в этом случае $f_1 < 0$. Если же первая линза собирательная, то это условие сводится к $l_{12} < f_1$. Эти результаты легче получить непосредственным геометрическим построением, что и рекомендуется сделать читателю.

Системы линз, аналогичные рассмотренной, применяются в современных ускорителях для фокусировки заряженных частиц. На них основан принцип так называемой *жесткой фокусировки.*

3. Каждая центрированная система может рассматриваться как сложная система, состоящая из нескольких подсистем. В качестве подсистем можно взять сферические границы раздела сред, на которых световые лучи испытывают преломление или отражение. Для сферической границы раздела коллинеарное соответствие выражается формулами (10.4). Из них и из формул (11.8) находим прежде всего: $x_H = x'_{H'} = 0$, т.е. обе главные плоскости совпадают между собой и проходят через точку пересечения рассматриваемой преломляющей поверхности с главной оптической осью системы. Для фокусных расстояний f и f' подсистем формулы (10.4) и (11.9) дают

$$f = \frac{Rn}{n' - n}, \qquad f' = -\frac{Rn'}{n' - n}. \tag{12.10}$$

Используя эти формулы, а также формулы (12.2), (12.3) и (12.4), можно рассчитать параметры любой центрированной системы.

4. В качестве примера проведем расчет параметров *толстой линзы*. Пусть R_1 и R_2 означают радиусы кривизны преломляющих сферических поверхностей линзы, n_1, n_2, n_3 — показатели преломления первой среды, вещества линзы и второй среды (рис. 48), f_1 и f_1' — фокусные расстояния при преломлении на передней поверхности линзы, f_2 и f_2' — на задней. В таком случае

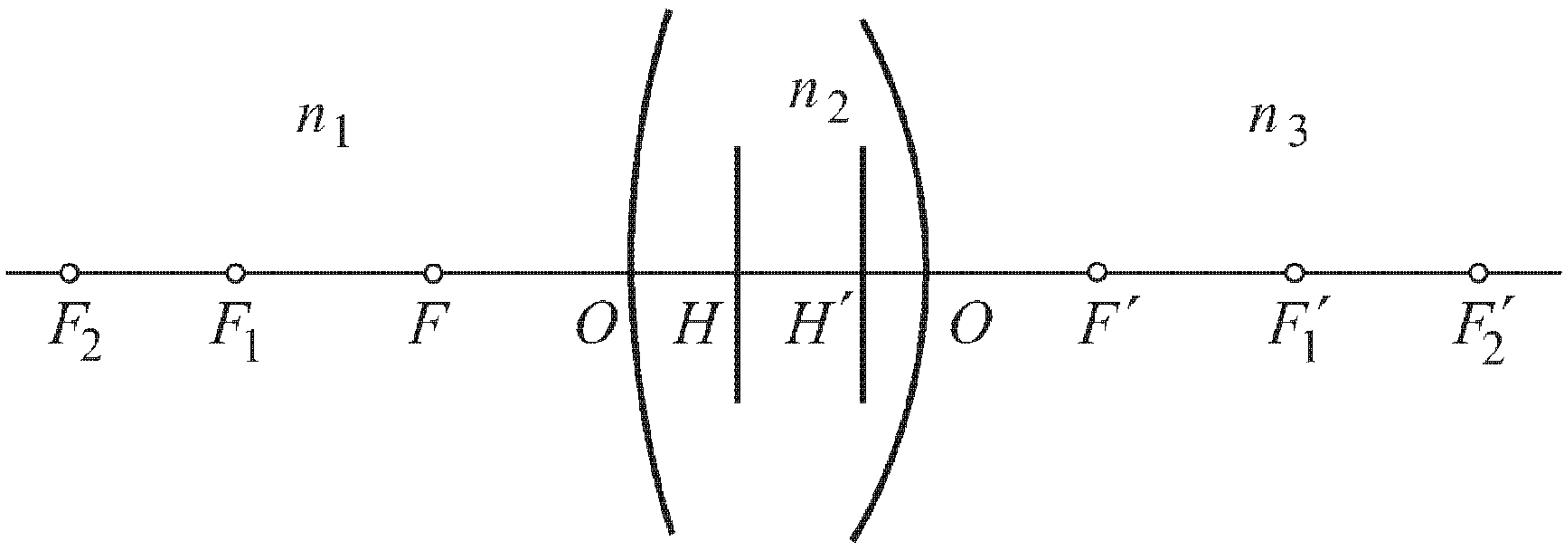


Рис. 48

$$f_1 = \frac{n_1 R_1}{n_2 - n_1}, \quad f_1' = -\frac{n_2 R_1}{n_2 - n_1},$$
$$f_2 = \frac{n_2 R_2}{n_3 - n_2}, \quad f_2' = -\frac{n_3 R_2}{n_3 - n_2}.$$

Будем рассматривать преломляющие поверхности линзы как центрированные подсистемы, а саму линзу — как сложную систему. Если d — толщина линзы, то $\Delta = f_1' + d - f_2$, или после подстановки значений f_1' и f_2'

$$\Delta = \frac{D}{(n_2 - n_1)(n_3 - n_2)}, \tag{12.11}$$

где введено обозначение

$$D = d(n_1 - n_2)(n_2 - n_3) + n_2[R_1(n_2 - n_3) + R_2(n_1 - n_2)]. \tag{12.12}$$

Для фокусных расстояний f и f' линзы из формул (12.3) получаем

$$f = -n_1 n_2 \frac{R_1 R_2}{D}, \quad f' = n_2 n_3 \frac{R_1 R_2}{D}. \tag{12.13}$$

Координаты фокальных точек линзы F и F' можно вычислить по формулам (12.2), которые дают

$$x_F = -n_1 n_2 \frac{n_2 - n_3}{n_1 - n_2} \frac{R_1^2}{D}, \quad x_{F'}' = n_2 n_3 \frac{n_1 - n_2}{n_2 - n_3} \frac{R_2^2}{D}. \tag{12.14}$$

При этом за начало координат в пространстве предметов линзы принят фокус F_1, а в пространстве изображений — фокус F_2'.

Найдем расстояния h и h' главных плоскостей линзы от точек O и O'. По определению

$$h = OH = OF_1 + F_1F + FH = -f_1 + x_F + f,$$
$$h' = O'H' = O'F_2' + F_2'F' + F'H' = -f_2' + x'_{F'} + f'.$$

Отсюда

$$h = n_1(n_2 - n_3)\frac{R_1 d}{D}, \quad h' = -n_3(n_1 - n_2)\frac{R_2 d}{D}. \tag{12.15}$$

Обычно показатели преломления крайних сред n_1 и n_3 одинаковы. Полагая в этом случае $n_1 = n_3 = 1$, $n_2 = n$, получим

$$f = -f' = -n\frac{R_1R_2}{D}, \tag{12.16}$$
$$h = (n-1)\frac{R_1 d}{D}, \quad h' = (n-1)\frac{R_2 d}{D}, \tag{12.17}$$
$$D = (n-1)[n(R_1 - R_2) - d(n-1)], \tag{12.18}$$
$$e = HH' = d - h + h'; \tag{12.19}$$

через e здесь обозначено расстояние главной плоскости H' от главной плоскости H (см. рис. 48).

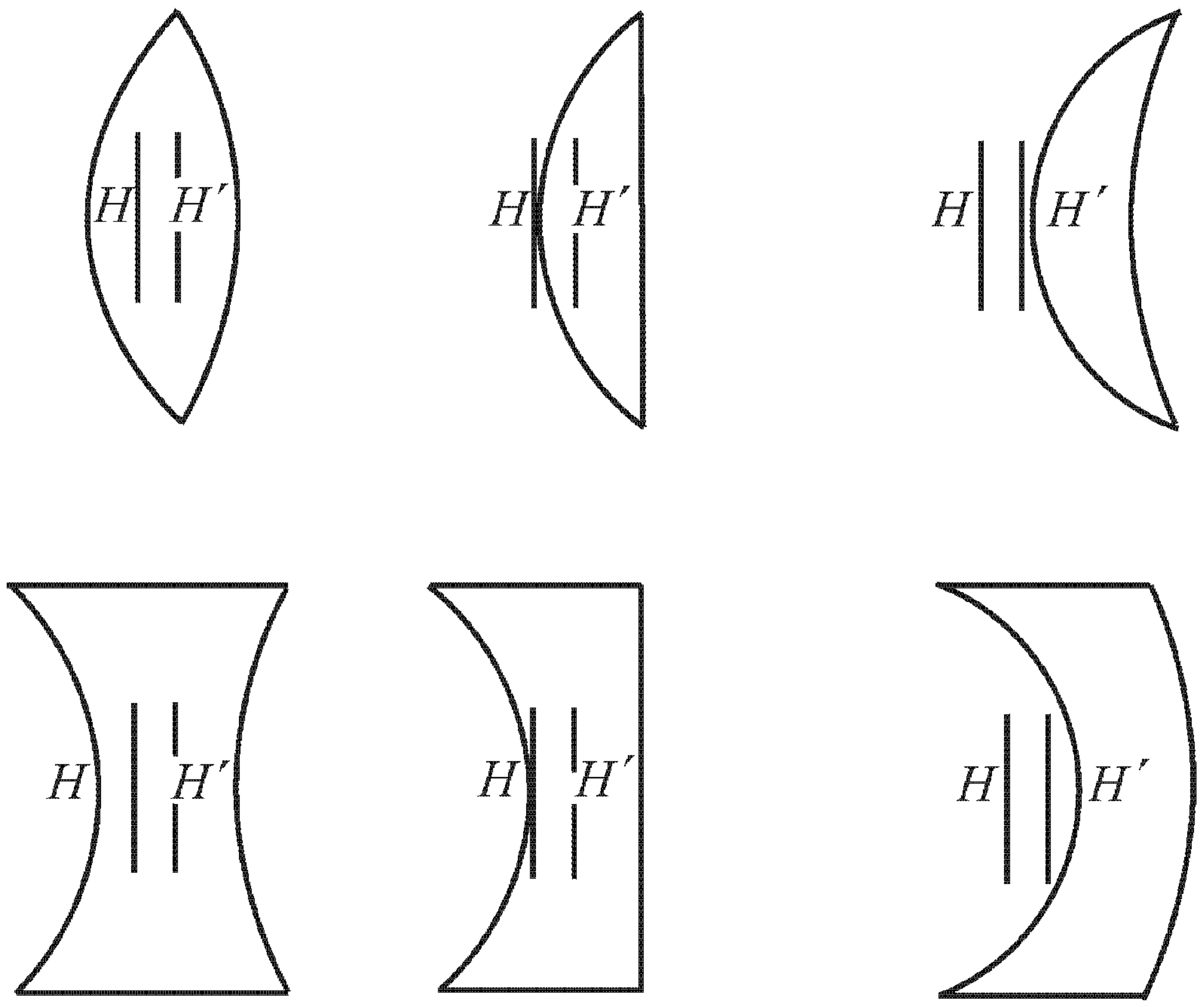


Рис. 49

Если линза не очень толстая, а разность $R_1 - R_2$ не слишком мала, то в выражении (12.18) слагаемым $d(n-1)$ можно пренебречь. В этом приближении

$$\frac{1}{f} = -\frac{1}{f'} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right), \tag{12.20}$$

$$h = \frac{R_1}{n(R_1 - R_2)}d, \quad h' = \frac{R_2}{n(R_1 - R_2)}d, \tag{12.21}$$

$$\frac{h}{h'} = \frac{R_1}{R_2}, \tag{12.22}$$

$$e = \frac{n-1}{n}d. \tag{12.23}$$

На рис. 49 изображены типичные линзы с указанием положений их главных плоскостей ($n = 1{,}5$).

ЗАДАЧИ

1. Для определения углового увеличения зрительной трубы *методом Рамсдена* (1735–1800) трубу устанавливают на бесконечность. Вывернув объектив, устанавливают на его место предмет определенной величины (экран с вырезом). Окуляр трубы дает действительное изображение взятого предмета. Пусть L — величина предмета, а l — величина его изображения. Показать, что угловое увеличение зрительной трубы равно L/l.

Решение. Примем за начала координатных систем фокальные точки окуляра. Тогда в формуле (11.17) следует положить $X = f_1'$, $f = f_2$. Это дает $L/l = f_1'/f_2$, т.е. увеличение трубы (см. § 11, п. 10).

2. Для определения фокусного расстояния собирательной линзы Бессель (1784–1846) предложил следующий метод.

С помощью линзы на экране получается действительное изображение предмета. Пусть A — расстояние от предмета до его изображения. Тогда $A = \xi' + e - \xi$. Исключая с помощью этого соотношения ξ' из (11.14), получим

$$\xi^2 + (A-e)\xi + (A-e)f = 0. \tag{12.24}$$

Если

$$A - e > 4f, \tag{12.25}$$

то уравнение (12.24) имеет два вещественных корня ξ_1 и ξ_2. В этом случае существуют два положения линзы, при которых на экране получаются действительные изображения предмета (при неизменном расстоянии между предметом и экраном). Чтобы перейти от одного изображения к другому, надо сместить линзу на расстояние $a = \xi_1 - \xi_2 = \sqrt{(A-e)^2 - 4f(A-e)}$, откуда

$$f = -f' = \frac{(A-e)^2 - a^2}{4(A-e)}. \tag{12.26}$$

Величины A и a можно измерить. Величина же e — расстояние между главными плоскостями — неизвестна. Для ее определения можно взять другое расстояние A_1 между предметом и экраном и измерить соответствующее смещение линзы a_1. Получится выражение вида (12.26), в котором A и a заменены на A_1 и a_1. Сравнивая эти два выражения, можно вычислить e. Для упрощения расчета можно пренебречь e^2 по сравнению с A^2. Это дает

$$f = -f' = \frac{A^2 - a^2}{4A} - \frac{A^2 + a^2}{4A^2}e. \tag{12.27}$$

§ 13. Ограничение лучей при помощи диафрагм

1. Четкие изображения, как правило, получаются только *в параксиальных лучах.* Непараксиальные лучи на практике устраняются *диафрагмами.* Роль диафрагм могут играть также оправы линз или зеркал.

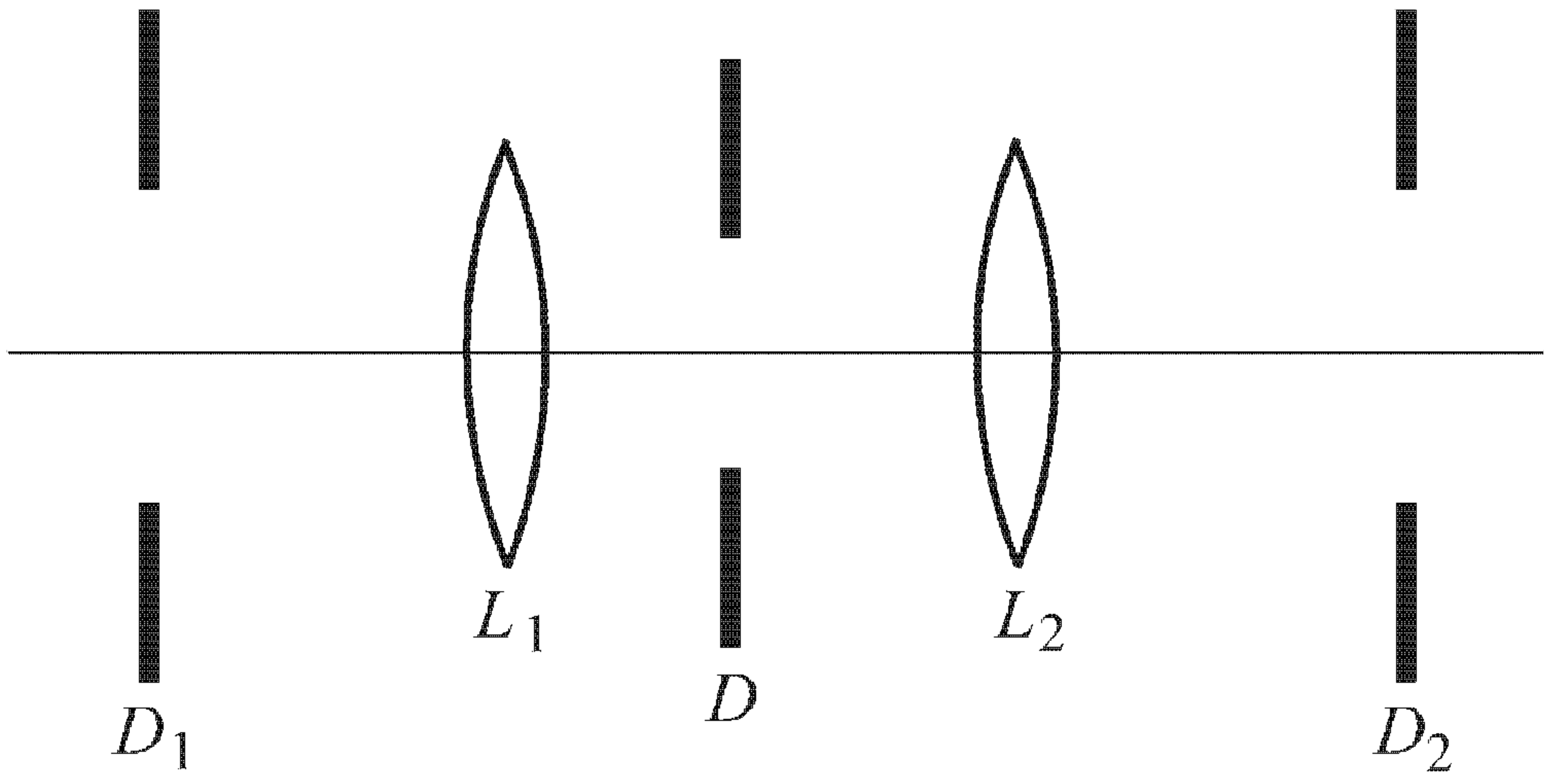


Рис. 50

Пусть D (рис. 50) — какая-либо диафрагма, а D_1 — ее изображение в параксиальных лучах впереди стоящими линзами. Если диафрагму D заменить диафрагмой D_1, то D_1 будет так же ограничивать параксиальные лучи, как и D. В самом деле, в приближении параксиальной оптики всякий луч, проходящий через край отверстия диафрагмы D_1, проходит и через край отверстия диафрагмы D. Следовательно, в этом приближении все реальные диафрагмы можно заменить их *изображениями,* получаемыми с помощью впереди стоящих линз. Тем самым все диафрагмы мысленно как бы переносятся в пространство предметов. Этим достигается то упрощение, что при исследовании действия диафрагм можно *отвлечься от преломления лучей.* Точно так же все реальные диафрагмы можно мысленно перенести в пространство изображений, заменив эти диафрагмы их изображениями, получаемыми с помощью позади стоящих линз.

2. Указанным способом мысленно перенесем все диафрагмы в пространство предметов (рис. 51). Воображаемая диафрагма DD', которая видна под наименьшим углом из точки предмета P, лежащей на главной оптической оси системы, всего более ограничивает лучи, исходящие из P. Она называется *входным зрачком* или *входным отверстием* системы. Реальная диафрагма, изображением которой является входной зрачок, носит название *апертурной* или *действующей диафрагмы.* Если апертурная диафрагма находится перед передней линзой, то она совпадает с входным зрачком. Таким образом, *апертурная диафрагма всего более диафрагмирует лучи, исходящие из точки предмета, лежащей на главной оптической оси системы.* Поэтому от размеров апертурной диафрагмы зависит *яркость изображения*. При изменении апертурной диафрагмы меняется *светосила прибора*. Угол $2u$, под которым виден входной зрачок из точки предмета P, называется *апертурным углом со стороны предмета*, или *углом раскрытия*. Его называют также *апертурой системы*. Изображение апертурной диафрагмы в параксиальных лучах, получаемое с помощью позади стоящих линз, называется *выходным зрачком* или *выходным отверстием* системы. Очевидно, выходной зрачок есть изображение входного зрачка в параксиальных лучах, даваемое всей системой. Выходной зрачок всего более диафрагмирует лучи, проведенные из точки P', являющейся изображением точки предмета P, которая лежит на главной оптической оси. Угол $2u'$, под которым выходной зрачок виден из P', называется *апертурным углом со стороны изображения*, или *углом проекции* системы (см. рис. 52).

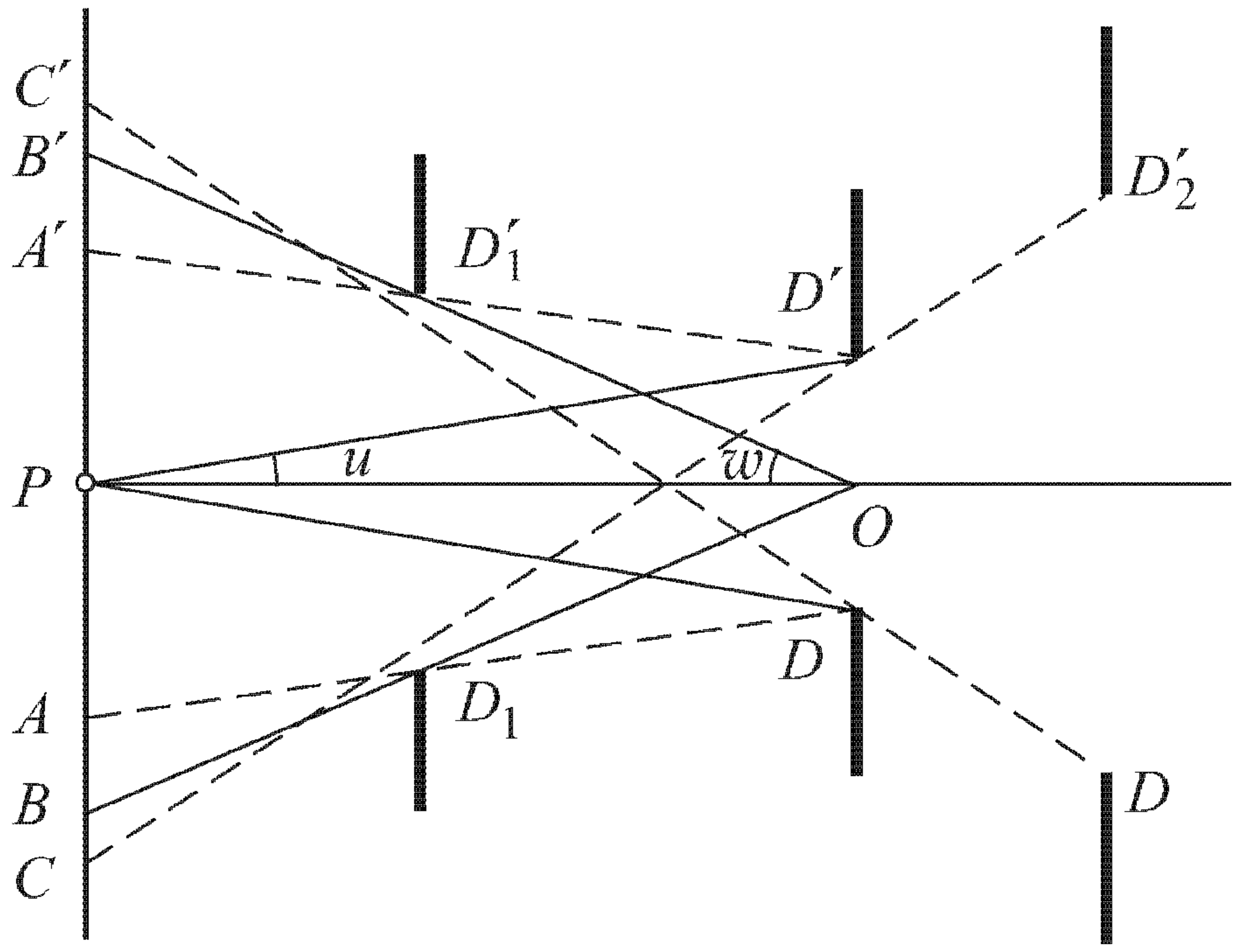


Рис. 51

В зрительных трубах апертурной диафрагмой часто служит край объектива. В таком случае этот край играет роль и входного зрачка, а его параксиальное изображение, даваемое окуляром, будет выходным зрачком. Если держать трубу на некотором расстоянии от глаза против светлого фона, то выходной зрачок будет виден как действительное или мнимое изображение.

В некоторых случаях апертурной диафрагмой может служить *радужная оболочка глаза*. Ее параксиальное изображение, даваемое роговой оболочкой и водянистым телом глаза, называется *зрачком глаза*.

Лучи, проходящие через центр апертурной диафрагмы, называются *главными лучами*. В приближении параксиальной оптики главный луч проходит также через центры входного и выходного зрачков.

Если предмет находится на бесконечности, то входным зрачком служит изображение той диафрагмы в пространстве предметов, которое имеет наименьший диаметр. Аналогично, если изображение получается в бесконечности, то выходным зрачком служит изображение той диафрагмы в пространстве изображений, которое имеет наименьший диаметр.

3. Если точка предмета лежит на главной оптической оси системы, то всякий луч, проходящий через входной зрачок, пройдет и через оптическую систему. Но если луч исходит из точки предмета, не лежащей на главной оптической оси, то этот луч может и не пройти через оптическую систему даже в том случае, когда он прошел через входной зрачок. Он может быть задержан другими диафрагмами. Чтобы разобрать этот вопрос, рассмотрим сначала только главные лучи.

Та диафрагма, которая всего больше ограничивает главные лучи, называется *диафрагмой поля зрения*. Если бы входной зрачок был бесконечно малым, то все лучи, проходящие через систему, могли бы считаться главными. Поэтому при бесконечно малом входном зрачке поле зрения было бы резко ограничено и определялось размерами диафрагмы поля зрения. Изображение диафрагмы поля зрения, даваемое впереди стоящими линзами, называется *входным окном*, или *входным люком* системы. Изображение диафрагмы поля зрения, даваемое сзади стоящими линзами, называется *выходным окном*, или *выходным люком*. Очевидно, выходной люк есть изображение входного люка, даваемое всей системой. На рис. 51 входной люк изображен диафрагмой D_1D_1'. При бесконечно малом входном зрачке угол поля зрения со стороны предмета равен углу $2w$, под которым входной люк виден из центра входного зрачка. Аналогично определяется угол поля зрения $2w'$ со стороны изображения.

В действительности входной зрачок имеет *конечные размеры*. Это может повести к нарушению резкости границы поля зрения. В самом деле, из рис. 51 видно, что любой луч, исходящий из точек, лежащих между A и A', и прошедший через входной зрачок, пройдет и через оптическую систему. Лучи из точки B' пройдут через оптическую систему, если они направлены в нижнюю половину входного зрачка; лучи, направленные в верхнюю половину входного зрачка, задерживаются диафрагмой поля зрения и через оптическую систему не проходят. Для лучей, выходящих из B', верхняя половина входного зрачка оказывается как бы закрытой диафрагмой поля зрения. Наконец, лучи, исходящие из точек, лежащих дальше C и C', через оптическую систему вообще не пройдут. Таким образом, на краю поля зрения будет наблюдаться непрерывное ослабление освещенности. Оно называется *затенением* или *виньетированием*. При наличии виньетирования поле зрения ограничено не резко. *Чтобы виньетирования не было, необходимо, чтобы входной люк системы лежал в плоскости предмета. Поэтому почти во всех визуальных оптических системах для устранения виньетиро-*

вания применяют диафрагмы поля зрения, помещаемые обычно в передней фокальной плоскости окуляра.

4. В оптических инструментах изображение обычно воспринимается на какую-нибудь поверхность, например матовое стекло, фотографическую пластинку, сетчатку глаза и т.п. Эта поверхность перпендикулярна к главной оптической оси системы и в первом приближении может быть принята за плоскость (*плоскость изображения*). Между тем в большинстве случаев предмет, изображение которого надо получить, бывает *пространственным*, а не представляет собой плоскости или какой-либо простой поверхности. Чему же соответствует картина, получаемая в плоскости изображения оптической системы?

Для ответа на этот вопрос найдем в пространстве предмета плоскость, оптически сопряженную с плоскостью изображения. Она называется *плоскостью установки* или *плоскостью наводки*. Спроецируем предмет из центра входного зрачка на плоскость установки. Эта проекция и будет тем объектом, изображение которого более или менее резко передает оптическая система. Действительно, главный луч, исходящий из любой точки предмета, проходит также через ее проекцию на плоскость установки. Он является центром пучка лучей, исходящих из этой точки. Если точка предмета лежит в плоскости установки, т.е. совпадает со своей проекцией, то ее изображение получится резким. Если же она не лежит в плоскости установки, то ее изображение получится в виде *кружка рассеяния*, центр которого является изображением проекции этой точки на плоскость установки. Чем больше апертурная диафрагма, тем шире пучки, исходящие из точек предмета, а следовательно, тем больше размеры соответствующих кружков рассеяния. Если бы пучки совсем не были ограничены диафрагмами или краями линз, то кружки рассеяния занимали бы всю плоскость изображения и изображение не могло бы вообще быть получено. Отсюда ясно, насколько важную роль играет диафрагмирование при получении изображений объемных предметов.

5. Для резкости изображения необходимо, чтобы диаметр кружка рассеяния d не превосходил известного предела (например, 0,1 мм). Этим определяется *глубина резко изображаемого пространства*. Найдем выражение для этой глубины в приближении параксиальной оптики. Пусть DD' (рис. 52) — входной, а D_1D_1' — выходной зрачки, P — точка предмета на оптической оси, P' — ее изображение, EE' — плоскость изображения. Опишем из P', как из центра, кружок диаметром d. Соединим диаметрально противоположные крайние точки A и A' этого кружка с краями D_1 и D_1' выходного зрачка. В пересечении с оптической осью получатся точки P_1' и P_2'. Пусть P_1 и P_2 — сопряженные им точки в пространстве предметов. Расстояние $|\delta X| = P_1P_2$ и определит глубину резко изображаемого пространства. Пусть X и X' — координаты сопряженных точек P и P' относительно главных фокусов системы. Из уравнения (11.16) с точностью до членов второго порядка получаем $X'\,\delta X + X\,\delta X' = 0$. С той же точностью, как видно из рис. 52, $\delta X' = d/\operatorname{tg} u' \approx d/u'$. Следовательно,

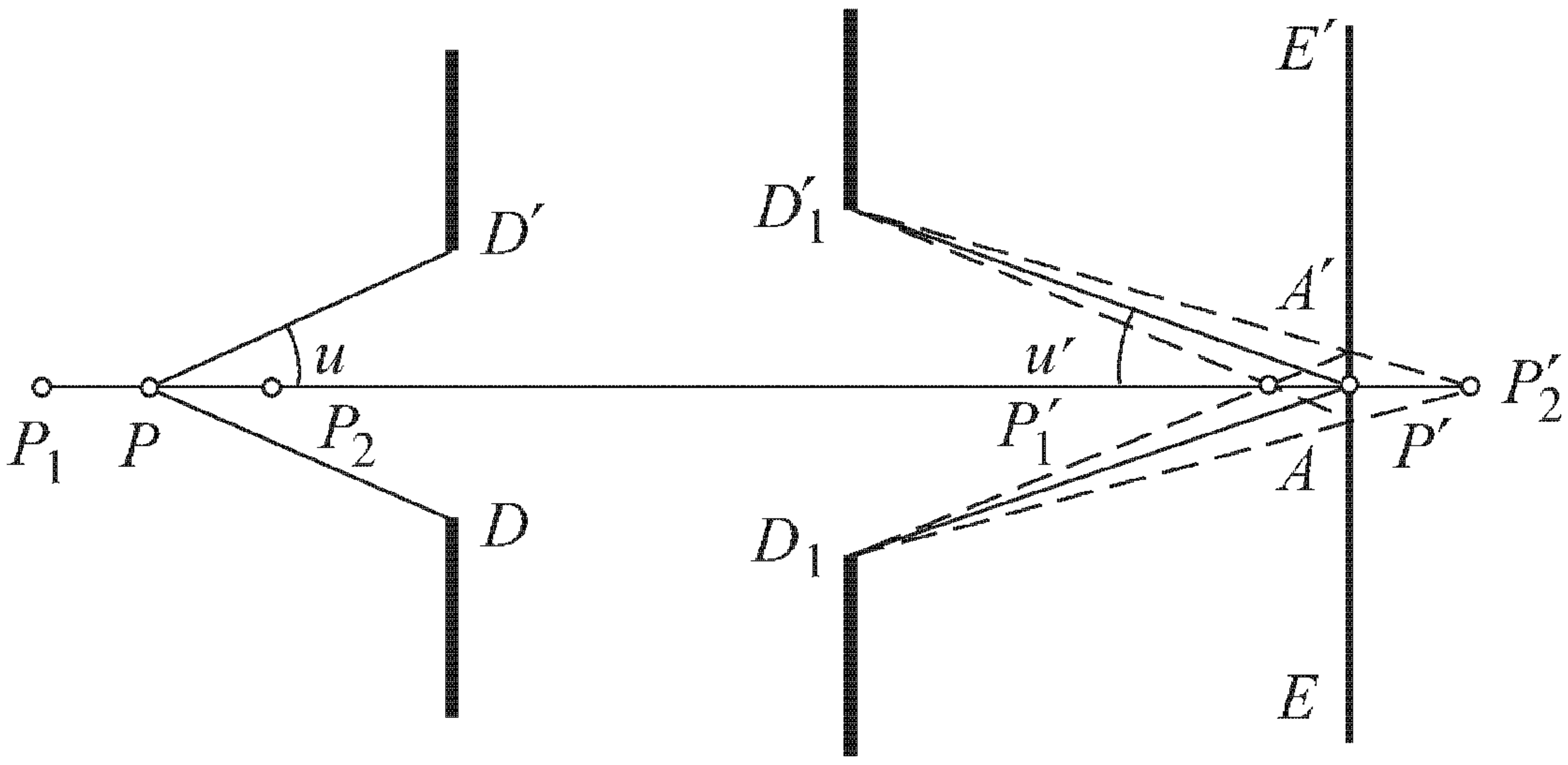


Рис. 52

$$|\delta X| = \left|\frac{X}{X'}\right| \frac{d}{u'}.$$

Комбинация этой формулы с формулами (11.12) и (11.17) приводит к результату

$$|\delta X| = \frac{|X|d}{fu}. \tag{13.1}$$

Если p — расстояние от предмета до входного зрачка, а r — радиус последнего, то $u = r/p$. Для объектива фотоаппарата можно принять $|X| = p$. Тогда

$$|\delta X| = \frac{p^2}{fr}d. \tag{13.2}$$

Каждому фотографу известно, что при недостатке глубины изображения надо либо уменьшить апертурную диафрагму, либо удалиться от фотографируемого объекта.

6. Положение входного зрачка по отношению к плоскости установки определяет *перспективу*, в которой виден предмет. Если входной зрачок расположен за предметом в направлении лучей, то при проецировании предмета из центра входного зрачка на плоскость установки ближние к оптической системе части предмета получат большее увеличение, чем дальние. При таком расположении близкие предметы на изображении будут получаться бóльшими, чем равные им более далекие предметы (*нормальная* или *энтоцентрическая перспектива*). Напротив, если входной зрачок лежит перед предметом, то на изображении близкие предметы получатся меньшими, чем равные им более далекие предметы (*гиперцентрическая перспектива*). Наконец, в промежуточном случае, когда входной зрачок удален в бесконечность, величина изображения не зависит от расстояния предмета до оптической системы (*телецентрическая перспектива*). Изменение расстояния до предмета сказывается в этом случае *не на величине изображения, а на его резкости.* Поэтому такая перспектива применяется в измерительных микроскопах, так как результат измерения при этом не зависит от точной установки измеряемого предмета. Чтобы ее осуществить, достаточно поместить апертурную диафрагму в задний фокус линзы L_1 (см. рис. 50).

В зрительных трубах, когда предмет практически находится в бесконечности, необходимо, чтобы величина изображения, проецирующегося на шкалу с делениями, не зависела от точной установки шкалы. Этого можно достигнуть, поместив апертурную диафрагму в передний фокус линзы L_2 (рис. 50). Тогда выходной зрачок удалится в бесконечность, главный луч будет параллелен оптической оси и встретит плоскость шкалы на одном и том же удалении от оптической оси. Следовательно, величина изображения на шкале не будет зависеть от ее положения.

§ 14. Астигматические пучки лучей. Каустика

1. При наличии непараксиальных лучей, а также при отсутствии осевой симметрии оптической системы (примером может служить цилиндрическая линза) сферическая волна, исходящая из светящейся точки, после прохождения через оптическую систему перестает быть сферической. В результате светящаяся точка уже не будет изображаться оптической системой в виде точки. Связанные с этим искажения оптических изображений называются *геометрическими* или *лучевыми аберрациями оптических систем.* Помимо лучевых существуют еще *хроматическая аберрация*, т.е. появление окрашенных каемок в изображении, когда оно получается в белом свете, а также *волновые* или *дифракционные аберрации*, обусловленные дифракцией света. Учет последних необходим при рассмотрении разрешающей способности оптических и спектральных приборов, а также других тонких вопросов, связанных с получением оптических изображений. Однако сначала мы отвлечемся от хроматической и дифракционных аберраций, предполагая, что световые пучки, формирующие изображение, — *монохроматические* и к ним применима *геометрическая оптика.*

2. Напомним сначала необходимые сведения из дифференциальной геометрии. Пусть S — какая-то гладкая поверхность, а O — произвольная точка на ней (рис. 53 *а*). Нормаль к поверхности S в точке O обозначим через $\mathbf{N}$. Проведем через $\mathbf{N}$ плоскость П, пересекающую поверхность S вдоль некоторой кривой L. Если плоскость П вращать вокруг нормали $\mathbf{N}$ в пределах 180°, то кривизна кривой L, вообще говоря, будет изменяться, достигая в каком-то положении L_1 максимума, а в другом положении L_2 — минимума. В дифференциальной геометрии доказывается, что нормальные сечения поверхности S максимальной и минимальной кривизны взаимно перпендикулярны. Эти сечения называются *главными нормальными сечениями поверхности* S, проходящими через точку O. Линии, по которым они пересекают поверхность S в окрестности точки O, называются *линиями кривизны*, а радиусы кривизны последних R_1 и R_2 — *главными радиусами кривизны поверхности* S. Если знаки R_1 и R_2 одинаковы, то линии кривизны обращены вогнутостями *в одну и ту же сторону*; если разные, то *в противоположные стороны*. Если в точке O $R_1 = R_2$, то эта точка называется *точкой закругления поверхности*.

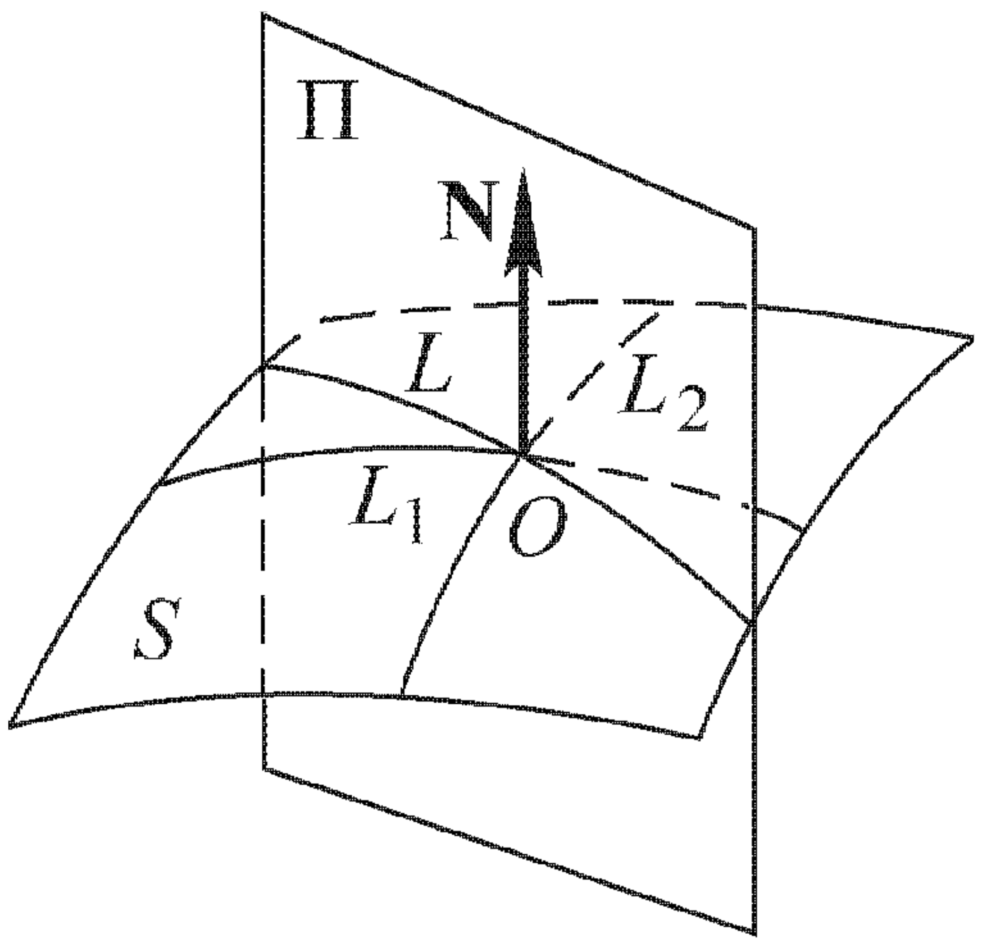


Рис. 53 *а*

3. Пусть теперь поверхность S является волновым фронтом. Возьмем на ней какую-то элементарную площадку dS. Бесконечно узкий пучок лучей волнового фронта, проходящий через эту площадку, называется *астигматическим*, если главные радиусы кривизны R_1 и R_2 не совпадают между собой. Луч, проходящий через центр площадки dS, называется *главным* [1]).

Вообще говоря, две нормали к поверхности S не пересекаются в одной точке. Однако нормали, восстановленные в бесконечно близких точках, лежащих на одной и той же линии кривизны, пересекаются в одной точке (если пренебречь величинами высших порядков малости). Заметив это, возьмем в качестве элемента dS бесконечно малую площадку $ABDC$, имеющую форму криволинейного четырехугольника, сторонами которого служат линии кривизны поверхности S (рис. 53 *б*). Лучи, лежащие в главных сечениях элемента волнового фронта dS, проходящих через его центр, сходятся в различных точках F_1 и F_2, называемых *фокальными точками*; расстояние F_1F_2 называется *астигматической разностью элементарного пучка*. Лучи, проходящие через противоположные стороны AB и CD, пересекутся в точках M_1 и N_1 соответственно. Лучи, исходящие из всякой промежуточной линии кривизны между сторонами AB и CD, сойдутся в какой-то промежуточной точке бесконечно малого отрезка M_1N_1, проходящего через фокальную точку F_1. Аналогично, лучи, исходящие из линий кривизны между сторонами BD и AC, пересекутся в какой-то точке другого бесконечно малого отрезка M_2N_2, проходящего через вторую фокальную точку F_2.

---

[1]) При отсутствии симметрии площадки понятие ее центра, строго говоря, теряет смысл. Однако и в этом случае можно указать какую-то, хотя и не совсем четко определенную, точку в середине площадки, которую можно рассматривать как центр последней.

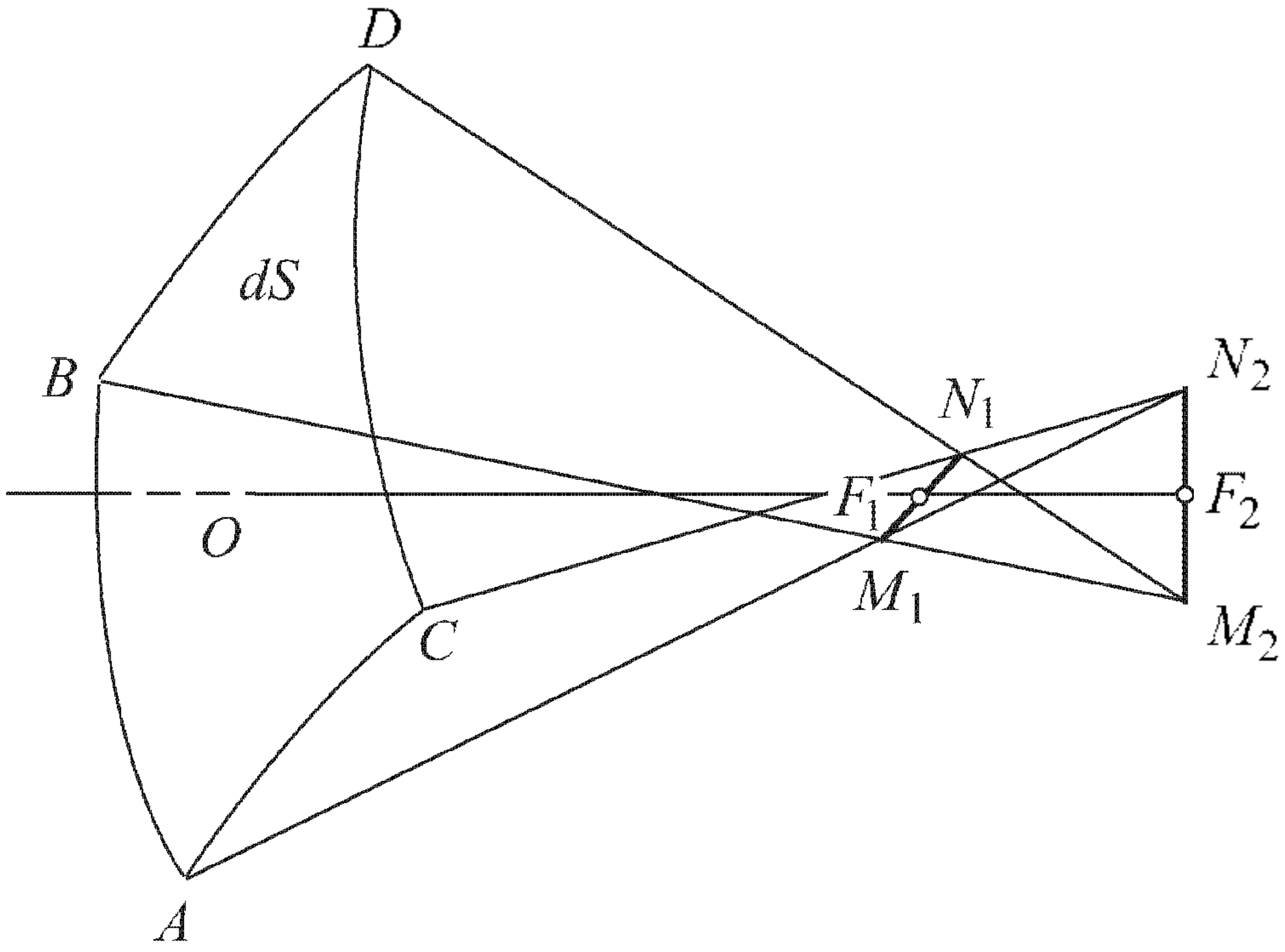


Рис. 53*б*

Отрезки M_1N_1 и M_2N_2 называются *фокальными отрезками* элементарного астигматического пучка лучей, исходящих от элемента волнового фронта $ABCD$. Если элементарный астигматический пучок лучей исходит из светящейся точки, то фокальные отрезки M_1N_1 и M_2N_2 принято называть *изображениями* этой точки, даваемыми *астигматическим пучком*, хотя точечного изображения в этом случае и не существует. Если элемент $ABCD$ обладает симметрией прямоугольника, то отрезки M_1N_1 и M_2N_2 будут перпендикулярны между собой, а также к главному лучу пучка. В общем случае это может быть и не так. При бесконечном уменьшении поперечных размеров пучка лучей отрезки M_1N_1 и M_2N_2 стягиваются в фокальные точки F_1 и F_2.

Таким образом, в отличие от гомоцентрического пучка, бесконечно узкий астигматический пучок дает не одно, а *два точечных изображения* F_1 и F_2 светящейся точки. Конечный пучок лучей можно разложить на элементарные астигматические пучки, каждому из которых соответствует пара фокальных точек. Геометрическое место этих точек есть двухлистная поверхность, называемая *каустической поверхностью*, или *каустикой*.

ЗАДАЧА

Точечный источник света P помещен в прозрачной однородной среде, ограниченной плоскостью. Световые лучи, исходящие из P, испытывают преломление на этой плоскости. Найти для них каустическую поверхность.

Решение. Каустика преломленных лучей состоит из двух листов. Один из них есть геометрическое место фокальных точек *меридиональных лучей*, т.е. лучей, лежащих в плоскости падения главного луча элементарного астигматического пучка. Другой — геометрическое место фокальных точек *экваториальных лучей*, т.е. лучей, лежащих в перпендикулярной плоскости, проходящей через главный луч элементарного пучка.

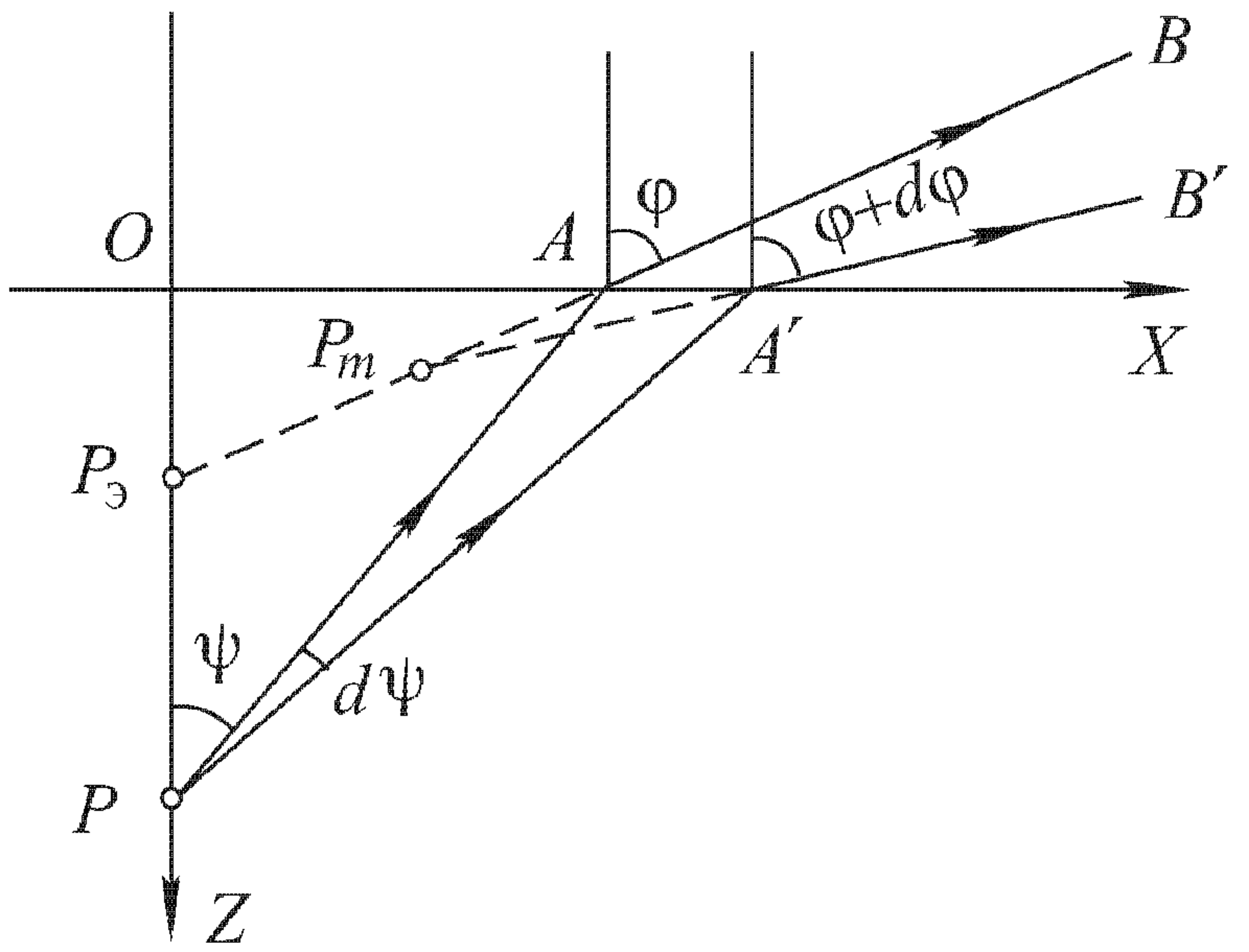


Рис. 54

Пусть n — показатель преломления среды, в которой помещен источник P. Показатель преломления пространства, с которым граничит эта среда, примем за единицу. Введем прямоугольную систему координат с началом O, расположенным на границе среды. За ось Z примем нормаль к поверхности среды, направив эту ось в сторону точки P (рис. 54). Ввиду осевой симметрии достаточно найти сечение каустической поверхности плоскостью, проходящей через ось Z. Прямую, вдоль которой эта плоскость пересекает границу среды, примем за ось X.

Найдем сначала каустику для меридиональных преломленных лучей. Уравнение преломленного луча AB будет

$$z = -\operatorname{ctg}\varphi(x - h\operatorname{tg}\psi),$$

где h — расстояние от точки P до границы среды, ψ — угол падения из среды на эту границу, φ — угол преломления. Для бесконечно близкого луча $PA'B'$ углы φ и ψ получат приращения $d\varphi$ и $d\psi$. Приращение координаты z при одном и том же значении абсциссы x при этом будет равно

$$dz = \frac{x - h\operatorname{tg}\psi}{\sin^2\varphi}\,d\varphi + \frac{h\operatorname{ctg}\varphi}{\cos^3\psi}\,d\psi,$$

или с использованием закона преломления $\sin\varphi = n\sin\psi$:

$$dz = \frac{x - h\operatorname{tg}\psi}{\sin^2\varphi}\,d\varphi + \frac{h\operatorname{ctg}\varphi\cos\varphi}{n\cos^3\psi}\,d\varphi.$$

Координаты x_m и z_m точки P_m, в которой пересекаются продолжения бесконечно близких меридиональных лучей AB и $A'B'$, найдутся отсюда, если приращение dz приравнять нулю. Это дает

$$x_m = h\left(\operatorname{tg}\psi - \frac{\cos^2\varphi\sin\varphi}{n\cos^3\psi}\right), \quad z_m = h\frac{\cos^3\varphi}{n\cos^3\psi}. \tag{14.1}$$

Это и есть *уравнение каустики для меридиональных лучей.* Используя его, нетрудно вывести формулу

$$l_m = l\frac{\cos^2\varphi}{n\cos^2\psi}, \tag{14.2}$$

где l — расстояние от предмета P до точки выхода A преломленного луча, а l_m — расстояние меридионального изображения P_m до той же точки.

Еще проще находится *каустика для экваториальных лучей.* Пусть PAB — один из лучей, исходящих из точки P (см. рис. 54). Если этот луч вращать вокруг перпендикуляра OP к преломляющей поверхности, то получится конус падающих и соответствующий ему конус преломленных лучей. Вершиной второго конуса будет точка $P_э$, в которой продолжение преломленного луча AB пересекает перпендикуляр PO. Бесконечно малые пучки падающих и преломленных лучей, лежащих на поверхностях указанных конусов, для которых луч PAB является главным, будут, очевидно, расположены в плоскостях, перпендикулярных к плоскости падения луча PAB. Значит, лучи этих пучков будут экваториальными, а точка $P_э$ — изображением в этих лучах. Таким образом, все фокальные точки экваториальных лучей расположатся на перпендикуляре PO, т.е. каустика таких лучей вырождится в отрезок этого перпендикуляра. Расстояние $l_э$ точки $P_э$ от точки выхода преломленного луча AB будет $P_эA = l\sin\psi/\sin\varphi$, т.е.

$$l_э = \frac{l}{n}. \tag{14.3}$$

Результаты вычислений представлены на рис. 55 для $n = 1{,}5$ (стекло). Каустика экваториальных лучей представляется вертикальным отрезком OP', длина которого равна h/n. Сечение каустики меридиональных лучей плоскостью рисунка есть кривая $BP'A$, для которой P' является *точкой возврата.* Продолжения преломленных лучей, изображенные на рисунке штриховыми прямыми, касаются кривой $BP'A$. Таким образом, если начертить еще сечение волнового фронта вышедших преломленных лучей, то

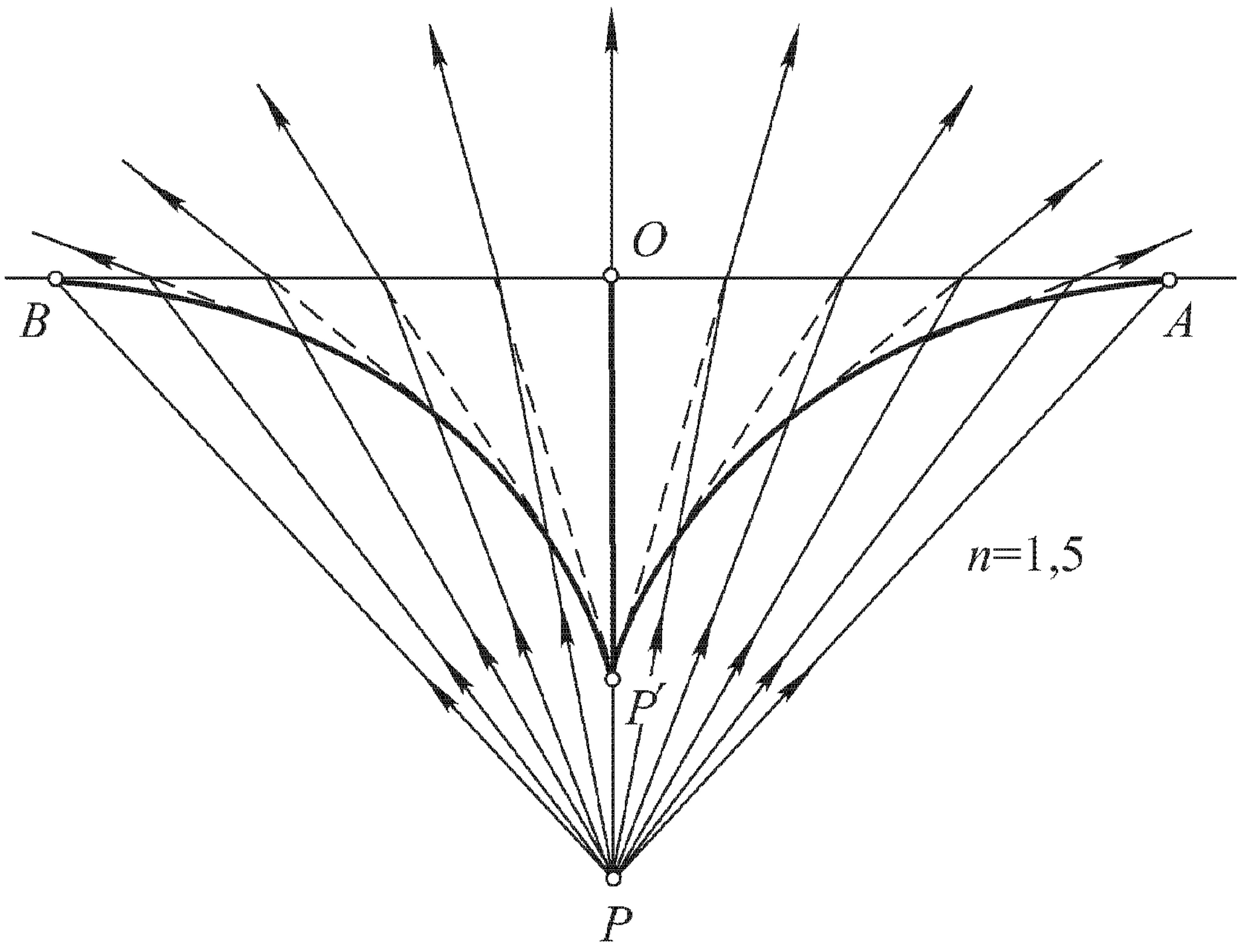


Рис. 55

для этого сечения кривая $BP'A$ будет *эволютой*. Сама каустика меридиональных лучей получится от вращения этой эволюты вокруг вертикальной прямой PO. Для параксиальных пучков фокальные точки меридиональных и экваториальных лучей совпадают и получаются в P'. Крайние точки A и B каустики меридиональных лучей получаются от пересечения предельных лучей полного отражения PA и PB с границей среды. Предлагаем читателю начертить каустику меридиональных и экваториальных лучей для случая $n < 1$.

§ 15. Геометрические аберрации центрированных систем

1. Геометрические аберрации в центрированных системах, т.е. отступления от параксиальной оптики, вызываются *непараксиальными пучками лучей*, участвующими в образовании оптических изображений. Дадим классификацию таких аберраций.

Произвольный луч в пространстве предметов можно задать, указав прямоугольные координаты x, z и η, ζ точек его пересечения с предметной плоскостью (т.е. плоскостью, проходящей через изображаемую точку P перпендикулярно к главной оптической оси) и плоскостью входного зрачка. После прохождения через оптическую систему луч пересечет плоскость параксиального изображения в точке с координатами y', z'. Координаты самого параксиального изображения (называемого в дальнейшем также *параксиальным фокусом*) обозначим через y'_0, z'_0. Тогда разности $\Delta y' = y' - y'_0$, $\Delta z' = z' - z'_0$ и можно принять за меру отступлений оптики реальной системы от предельного случая параксиальной оптики.

Координаты y', z' будут функциями аргументов y, z, η, ζ:

$$y' = f_1(y, z, \eta, \zeta), \qquad z' = f_2(y, z, \eta, \zeta).$$

Для классификации геометрических аберраций разложим эти функции в степенные ряды по своим аргументам. Линейные члены этих разложений, пропорциональные y и z, соответствуют параксиальной оптике. Линейные члены по η и ζ не войдут, так как в параксиальном приближении y' и z' не зависят от наклона лучей, выходящих из точки P. Не могут войти и члены четных степеней ввиду осевой симметрии оптической системы. Из всего этого следует, что разложения в степенные ряды отклонений $\Delta y' = y' - y'_0$ и $\Delta z' = z' - z'_0$ могут содержать только члены *нечетных степеней* по y, z, η, ζ, причем эти разложения могут начаться с членов, степень которых *не ниже трех*. Считая аргументы y, z, η, ζ малыми, сохраним в разложениях только члены третьей степени. Аберрации, вычисленные в этом приближении, называются *первичными*, или *аберрациями третьего порядка*.

Члены пятой степени вызывают аберрации *пятого порядка*, и т.д. Мы ограничимся только аберрациями третьего порядка.

Переходя к векторным обозначениям, введем три вектора, перпендикулярных к главной оптической оси системы:

$$\mathbf{r} = y\mathbf{j} + z\mathbf{k}, \quad \mathbf{r}' = y'\mathbf{j} + z'\mathbf{k}, \quad \boldsymbol{\sigma} = \eta\mathbf{j} + \zeta\mathbf{k},$$

где $\mathbf{j}$ и $\mathbf{k}$ — единичные векторы, указывающие направления координатных осей Y и Z. Вектор $\mathbf{r}$ определяет положение точки-предмета P в предметной плоскости, вектор $\mathbf{r}'$ — точки пересечения выходящего луча с параксиальной плоскостью изображения, вектор $\boldsymbol{\sigma}$ — точки пересечения падающего луча с плоскостью входного зрачка. Вектор $\Delta\mathbf{r}' = \Delta y'\mathbf{j} + \Delta z'\mathbf{k}$ можно разложить по векторам $\boldsymbol{\sigma}$ и $\mathbf{r}$. Ввиду осевой симметрии коэффициенты этих разложений могут зависеть только от «инвариантов вращения» $\boldsymbol{\sigma}^2$, $(\boldsymbol{\sigma}\mathbf{r})$ и $\mathbf{r}^2$. Поэтому с точностью до членов третьей степени включительно

$$\Delta\mathbf{r}' = [A\boldsymbol{\sigma}^2 + B(\boldsymbol{\sigma}\mathbf{r}) + C\mathbf{r}^2]\boldsymbol{\sigma} + [D\boldsymbol{\sigma}^2 + E(\boldsymbol{\sigma}\mathbf{r}) + F\mathbf{r}^2]\mathbf{r}, \qquad (15.1)$$

где A, B, C, D, E, F — постоянные коэффициенты, зависящие от устройства оптической системы и от положения предметной плоскости.

В дальнейшем будем понимать под σ радиус входного зрачка, выделяя тем самым падающие лучи (или их продолжения), проходящие через точки окружности входного зрачка. Тогда вектор $\Delta\mathbf{r}'$ окажется разложенным по степеням радиуса входного зрачка. Назовем *аберрационной кривой* кривую, по которой плоскость параксиального изображения пересекает пучок лучей, проведенных из точки-объекта P через окружность входного зрачка. Изображением точки P в параксиальной плоскости изображения будет не точка, а какое-то пятнышко, ограниченное аберрационной кривой. Для наглядности можно представить, что в качестве апертурной взята ирисовая диафрагма, радиус которой можно непрерывно менять. Тогда разложение (15.1) определит, как в рассматриваемом приближении будет меняться аберрационная кривая при изменении радиуса этой диафрагмы. Отступления от параксиальной оптики определяются, конечно, суммой (15.1) в целом, а не отдельными слагаемыми, из которых она состоит. Однако при классификации аберраций имеет смысл рассматривать каждое слагаемое в отдельности и рассуждать так, как если бы остальных слагаемых не было совсем. Тогда, в зависимости от степени σ, все аберрации третьего порядка можно разбить на четыре группы, которые мы и рассмотрим.

2. Сферическая аберрация. Эта аберрация вызывается членом третьей степени $A\sigma^2\boldsymbol{\sigma}$, так что при наличии одной только сферической аберрации $|\Delta\mathbf{r}'| = A\sigma^3 = \text{const}$.

Следовательно, аберрационной кривой будет *окружность* с центром в параксиальном фокусе и с радиусом $A\sigma^3$. Каждая точка будет изображаться в виде кружка рассеяния, радиус которого пропорционален кубу радиуса входного зрачка и не зависит от положения этой точки. Освещенность кружка рассеяния быстро убывает от центра к краям.

Происхождение сферической аберрации наглядно пояснено на рис. 56. Пусть точечный объект лежит на главной оптической оси системы. Выходящие из него параксиальные лучи встречают плоскость параксиальных изображений в точке G. Лучи, проходящие через окружность выходного зрачка DD', сойдутся на оптической оси в точке M, которая может лежать как ближе, так и дальше G. Лучи, проходящие через какую-либо окружность в плоскости выходного зрачка, концентрическую с окружностью DD', сойдутся на оптической оси между точками M и G. Расстояние MG называется *продольной сферической аберрацией.* Если в плоскости параксиальных изображений AA' поместить экран, то на нем получится светлый кружок радиуса GA. Радиус GA называется *поперечной сферической аберрацией.* С точностью до членов третьего порядка включительно поперечная аберрация пропорциональна кубу апертуры $2u$. Отсюда следует, что продольная аберрация пропорциональна квадрату апертуры.

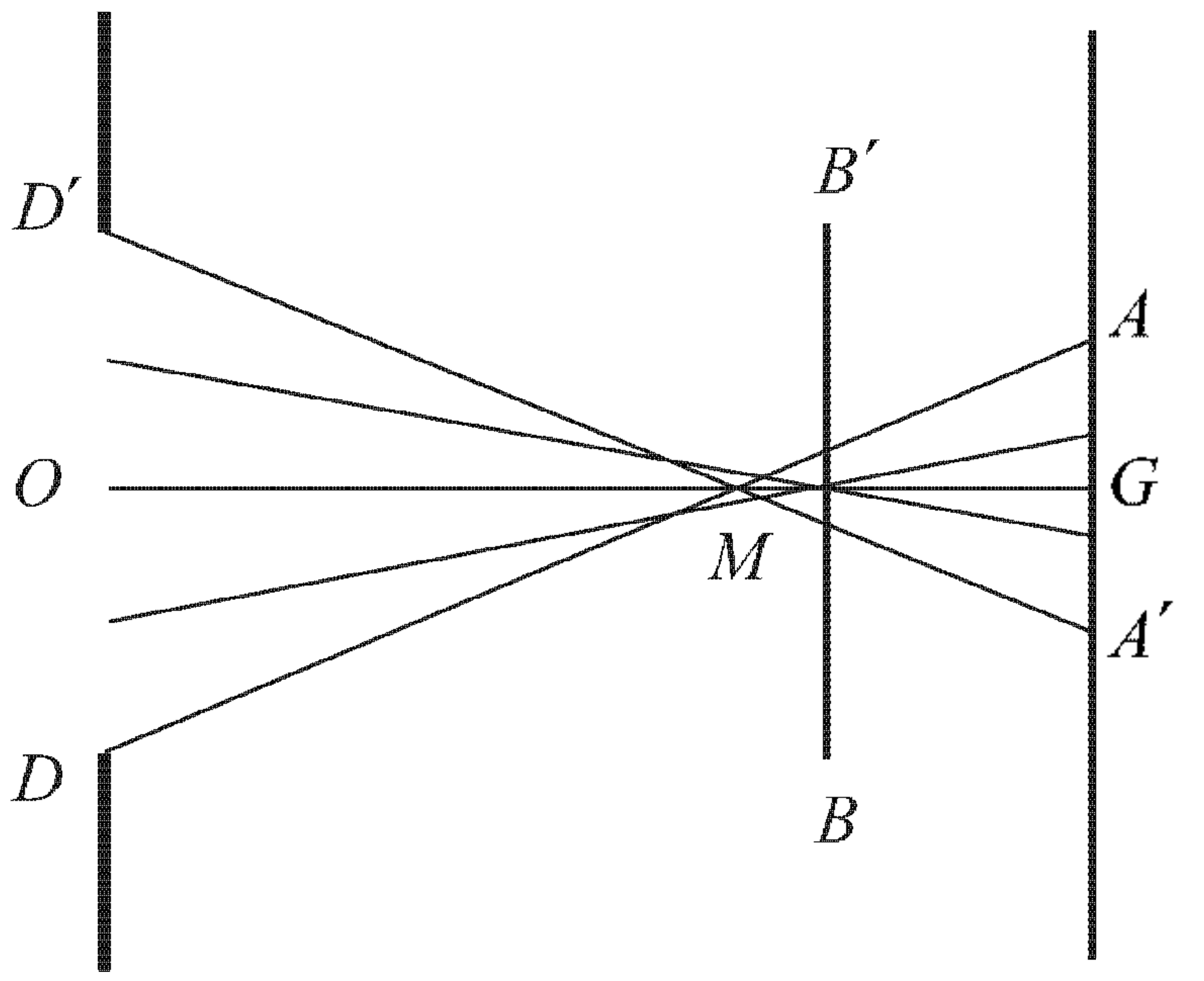


Рис. 56

Если экран перемещать от плоскости AA' по направлению к M, то радиус кружка рассеяния сначала будет уменьшаться, а затем начнет увеличиваться. Нетрудно показать, что наименьший кружок рассеяния получится, когда экран займет положение BB' на расстоянии ${}^3\!/_4 MG$ от плоскости параксиальных изображений AA'. Однако плоскость BB', строго говоря, не будет плоскостью наилучшей отчетливости изображения. При нахождении последней необходимо учитывать не только размеры кружка рассеяния, но и *распределение освещенности* внутри этого кружка. Исходя из дифракционных соображений, можно показать, что при наличии одной только сферической аберрации плоскость наилучшей отчетливости изображения проходит посередине между точками M и G.

Сферическая аберрация есть единственная из геометрических аберраций, остающаяся и в том случае, когда точка-объект находится на главной оптической оси системы. Все прочие геометрические аберрации в этом случае исчезают. Это верно в любом приближении, а не только для аберраций третьего порядка.

3. Кома. Если в разложении (15.1) отличны от нуля только коэффициенты B и D, то соответствующая аберрация называется *комой.* В этом случае

$$\Delta\mathbf{r}' = B(\boldsymbol{\sigma}\mathbf{r})\boldsymbol{\sigma} + D\sigma^2\mathbf{r}.$$

Отсюда легко получить

$$\left[\Delta\mathbf{r}' - \left(D + \frac{B}{2}\right)\sigma^2\mathbf{r}\right]^2 = \left(\frac{B}{2}\,\sigma^2\mathbf{r}\right)^2.$$

Следовательно, аберрационной кривой будет *окружность* радиуса $^1\!/_2B\sigma^2r$, центр которой смещен от параксиального фокуса в направлении вектора $\mathbf{r}$ на расстояние $(D + {}^1\!/_2B)\sigma^2r$. Теперь легко получить представление о характере изображения точечного объекта при наличии одной только комы в отсутствие других аберраций. Для этого проведем в плоскости входного зрачка произвольную окружность, центр которой совпадает с центром зрачка. Лучи, исходящие из точечного объекта и проходящие через эту окружность, пересекут плоскость параксиального изображения также по окружности. Совокупность таких окружностей и даст изображение рассматриваемого точечного объекта в этой плоскости. Окружности имеют две прямолинейные огибающие, пересекающиеся в параксиальном фокусе и составляющие между собой угол α, определяемый соотношением $\sin(\alpha/2) = B/(B + 2D)$. Более подробное исследование, которое мы не приводим, показывает, что $B = 2D$, а потому $\alpha = 60°$. Направление вектора $\mathbf{r}$ является биссектрисой угла между огибающими (рис. 57). Изображение точки, таким образом, напоминает комету. Отсюда и произошло название «кома».

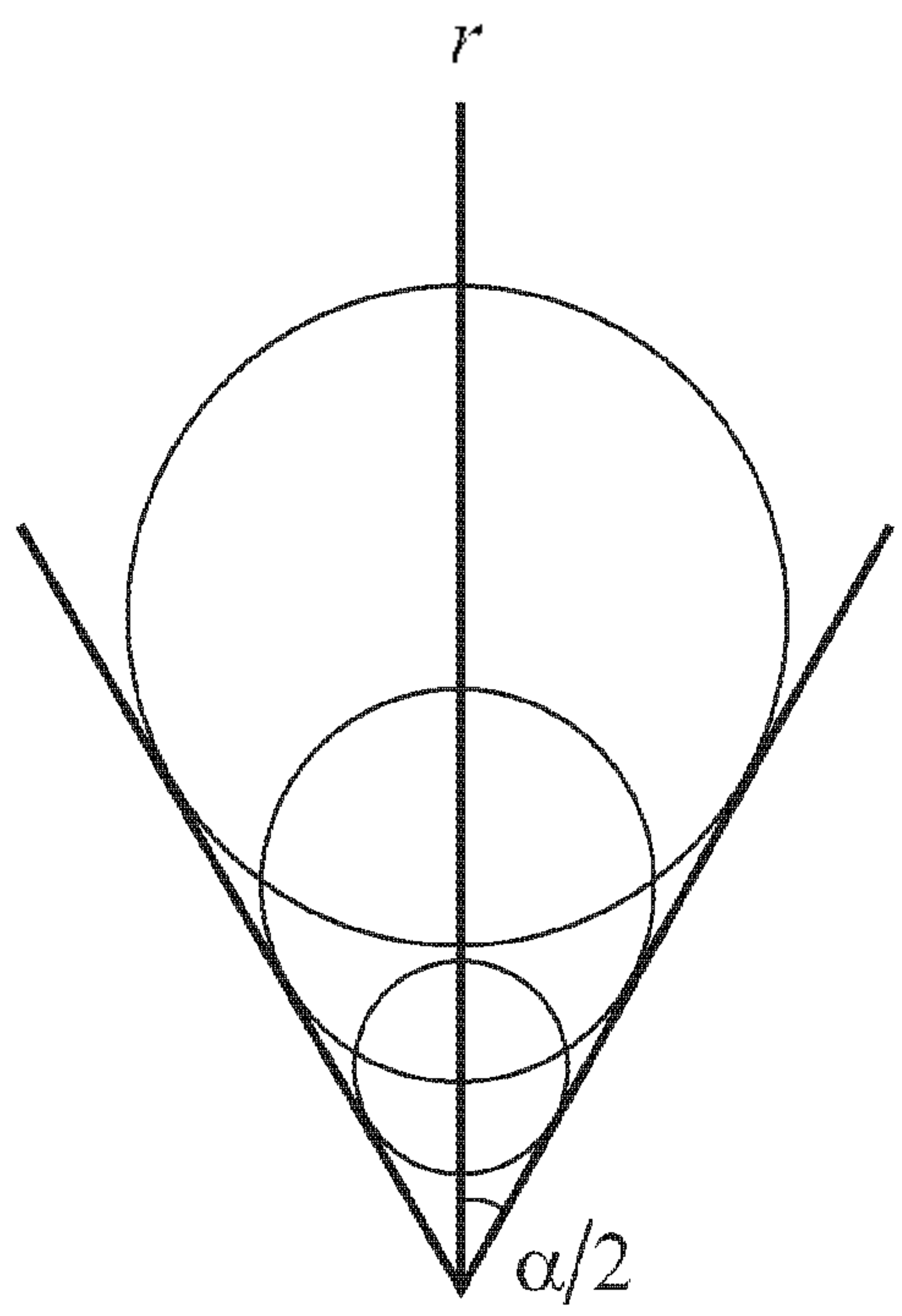


Рис. 57

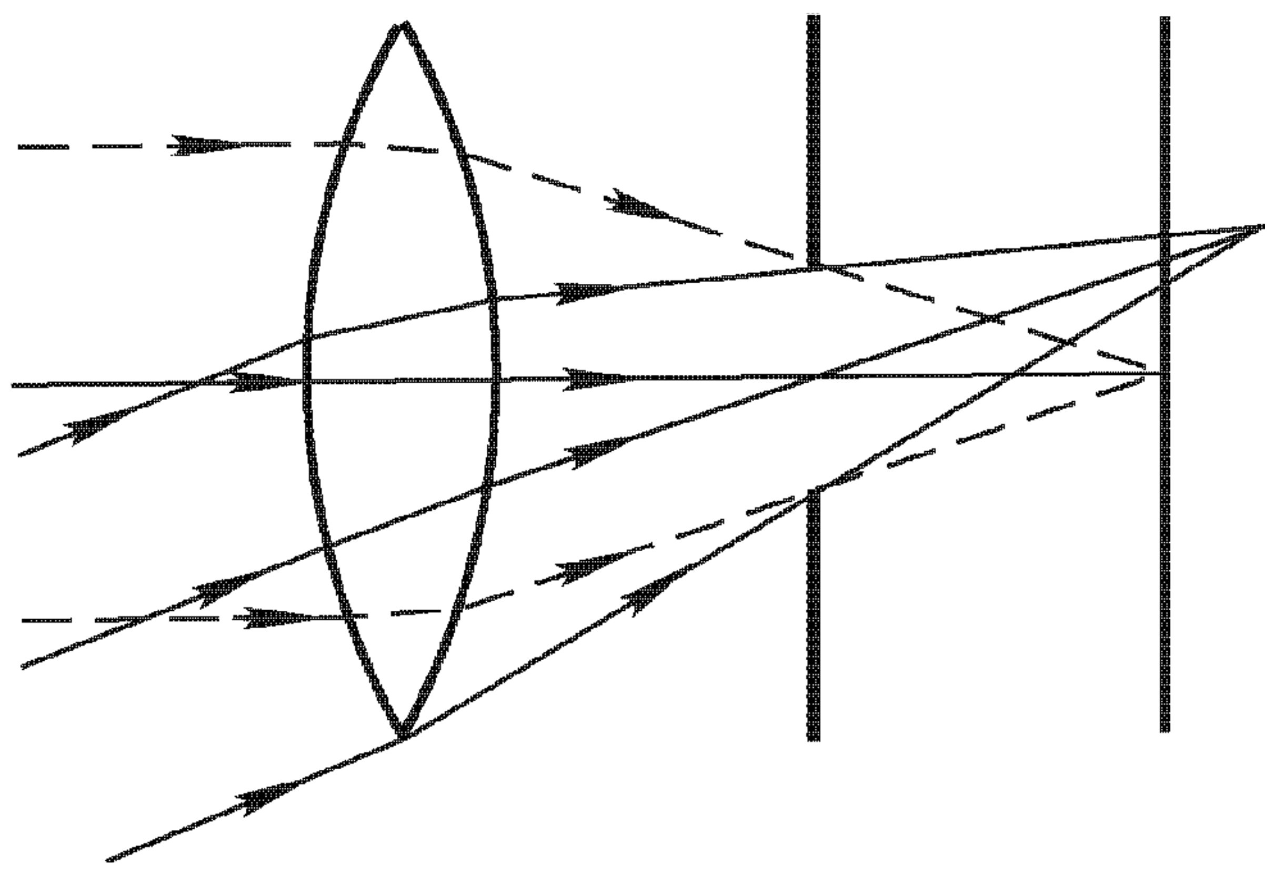


Рис. 58

Происхождение комы ясно из рис. 58. Она обусловлена косыми пучками лучей.

4. Астигматизм косых пучков и искривление плоскости изображения. Эти аберрации удобно рассматривать совместно, так как обе они обусловлены членами первой степени по $\boldsymbol{\sigma}$ и второй степени по $\mathbf{r}$. Они возникают, когда оба коэффициента C и E или один из них отличны от нуля. Если все прочие коэффициенты равны нулю, то формула (15.1) переходит в

$$\Delta\mathbf{r}' = Cr^2\boldsymbol{\sigma} + E(\boldsymbol{\sigma}\mathbf{r})\mathbf{r}. \tag{15.2}$$

Для определения формы аберрационной кривой ось Y проведем через точку-объект P. Тогда $\mathbf{r} = y\mathbf{j}$. Уравнение окружности входного зрачка запишем в параметрической форме $\eta = \sigma\cos\varphi$, $\zeta = \sigma\sin\varphi$, где φ — центральный угол, рассматриваемый как параметр. Таким путем из предыдущего соотношения получаем уравнение аберрационной кривой:

$$\begin{aligned} \Delta y' &= (C+E)y^2\eta = (C+E)\sigma y^2\cos\varphi, \\ \Delta z' &= Cy^2\zeta = C\sigma y^2\sin\varphi. \end{aligned} \tag{15.3}$$

Это — эллипс, центр которого находится в параксиальном фокусе, оси параллельны координатным осям Y и Z, а их длины пропорциональны радиусу входного зрачка и квадрату расстояния изображаемой точки от главной оптической оси. Изображением точки будет светлое пятнышко, ограниченное аберрационной кривой. Это указывает на то, что пучок лучей, дающий изображение, — астигматический. При параллельном смещении экрана, на котором получается изображение, вдоль оптической оси оно по-прежнему сохраняет форму эллипса, но форма и размеры эллипса изменяются. При двух положениях экрана эллипс вырождается в прямолинейные отрезки, один из которых параллелен оси Y, а другой — оси Z.

Для доказательства поместим начало координат в центре выходного зрачка и будем характеризовать луч в пространстве изображений точками пересечения его с плоскостью выходного зрачка и с плоскостью параксиального изображения. Координаты первой точки будут $(0, \eta', \zeta')$, второй — $(x', y'+\Delta y', \Delta z')$. Если X, Y, Z — текущие координаты, то уравнение рассматриваемого луча запишется в виде

$$Y = \eta' + \frac{y' + \Delta y' - \eta'}{x'}X, \qquad Z = \zeta' + \frac{\Delta z' - \zeta'}{x'}X.$$

Полагая здесь $\eta' = \zeta' = \Delta y' = \Delta z' = 0$, найдем поперечные координаты параксиального фокуса: $Y_0 = y'X/x'$, $Z_0 = 0$ (продольная координата того же фокуса равна x'). Таким образом,

$$\begin{aligned} Y - Y_0 &= \left(1 - \frac{X}{x'}\right)\eta' + \frac{X}{x'}\Delta y', \\ Z - Z_0 &= \left(1 - \frac{X}{x'}\right)\zeta' + \frac{X}{x'}\Delta z'. \end{aligned} \tag{15.4}$$

В последних слагаемых в окрестности параксиального фокуса можно считать $X = x'$, так как это вносит ошибки высшего порядка малости, не учитываемые в рассматриваемом приближении. По той же причине в формулах (15.3) можно положить $y = y'/\beta_1$, $z = z'/\beta_1$, $\eta = \eta'/\beta_2$, $\zeta = \zeta'/\beta_2$, где β_1 и β_2 — линейные поперечные увеличения для объектов, лежащих соответственно в предметной плоскости и в плоскости входного зрачка, как они выражаются в параксиальном приближении. Учтя еще, что y и Y являются лишь различными обозначениями одной и той же ординаты, можем записать (15.3) в виде

$$\Delta y' = \frac{C' + E'}{x'}\eta' Y^2, \quad \Delta z' = \frac{C'}{x'}\zeta' Y^2,$$

где C' и E' — новые постоянные, а абсцисса x' введена в знаменатели для удобства. В результате всего этого формулы (15.4) в окрестности параксиального фокуса преобразуются в

$$\begin{aligned} Y - Y_0 &= \frac{\eta'}{x'}\left[x' - X + (C' + E')Y^2\right], \\ Z - Z_0 &= \frac{\zeta'}{x'}\left[x' - X + C'Y^2\right]. \end{aligned} \tag{15.5}$$

Если

$$X - x' = (C' + E')Y^2, \tag{15.6}$$

то $Y - Y_0 = 0$, т.е. эллипс вырождается в прямолинейный отрезок, параллельный оси Z. Это есть *фокальный отрезок*, образованный *меридиональными лучами*. Аналогично, если

$$X - x' = C'Y^2, \tag{15.7}$$

то $Z - Z_0 = 0$, и эллипс переходит в другой фокальный отрезок, параллельный оси Y. Он образуется *экваториальными лучами*. От вращения кривых (15.6) и (15.7) вокруг главной оптической оси получаются две поверхности, касающиеся друг друга в общей точке пересечения их с главной оптической осью. Эти две поверхности и образуют каустику лучей, прошедших через оптическую систему. Вообще говоря, они имеют только одну общую точку. Каустика и есть та поверхность, в которую переходит плоскость изображения параксиальной оптики. Таким образом, имеет место не только астигматизм, но и *искривление поверхности изображения*.

5. Дисторсия. Коэффициент F не равен нулю, все прочие коэффициенты равны нулю. Поэтому из (15.1) получаем

$$\Delta r' = Fr^3.$$

Отсюда видно, что при наличии только рассматриваемой аберрации каждая точка изображается резко в виде точки, каковы бы ни были размеры диафрагмы. Однако отклонение изображения точки от соответствующего параксиального фокуса пропорционально кубу ее расстояния r от главной оптической оси. Поэтому происходит *искажение* (*дисторсия*) изображения. Прямые линии, проходящие через главную оптическую ось, изображаются в виде прямых. Все прочие прямые при изображении искривляются. При положительном F изображения точек смещаются относительно соответствующих параксиальных фокусов *наружу*, т.е. от главной оптической оси. Такая дисторсия называется *подушкообразной* (рис. 59 *б*). При отрицательном F смещения происходят внутрь — к главной оптической оси. Соответствующая дисторсия называется *бочкообразной* (рис. 59 *в*).

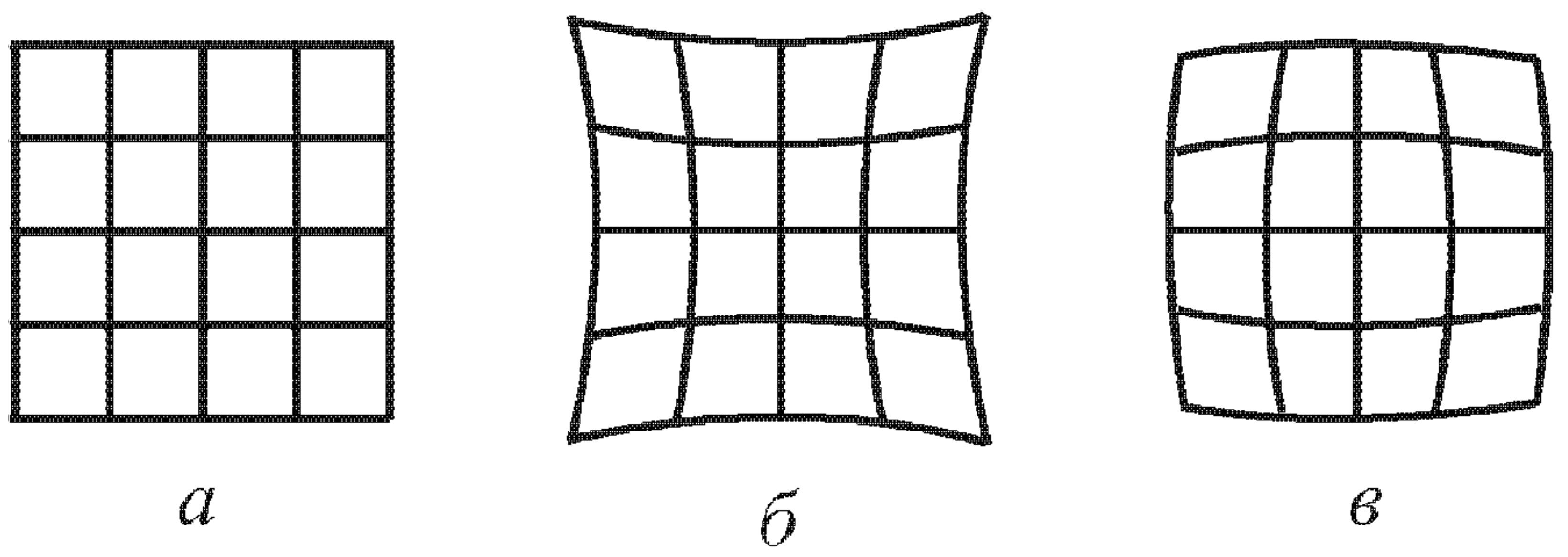


Рис. 59

Все эти выводы сохраняются в силе и для *аберраций высших порядков*. В общем случае дисторсия обусловлена всеми членами разложения $\Delta r'$, не содержащими степеней радиуса входного зрачка σ. В любом приближении при наличии одной только дисторсии точка всегда изображается в виде точки, каковы бы ни были размеры диафрагмы.

6. Устранение (точнее — ослабление) геометрических аберраций на практике достигается комбинацией различных линз (и зеркал), изготовляемых из оптического стекла с различными оптическими характеристиками. Полное устранение всех аберраций невозможно. На практике речь может идти только об определенном оптимуме, зависящем от задач, которые должен решать оптический прибор. Наиболее вредными из всех аберраций третьего порядка являются обычно сферическая аберрация и кома. В большинстве случаев их надо как можно больше ослабить. Уменьшая диафрагму, можно практически полностью устранить обе эти аберрации. После этого подбором линз надо устранить дисторсию, а затем астигматизм и изгиб поверхности изображения. Но уменьшение диафрагмы уменьшает яркость изображения и увеличивает дифракционные ошибки. Дисторсия вредна в фотографических объективах. В астрономических приборах ее можно допускать, так как она не влияет на резкость изображения, а вызывает только искажение его, которое можно учесть вычислением.

§ 16. Хроматическая аберрация

1. Если используется белый свет, то в изображении возникают дополнительные аберрации. Действительно, показатель преломления зависит от длины волны (дисперсия света). Поэтому оптическая система дает не одно, а множество монохроматических изображений, отличающихся друг от друга по величине и положению. В этом можно убедиться, разложив белый свет на монохроматические составляющие и воспользовавшись принципом суперпозиции. Результирующее изображение, получающееся от наложения таких монохроматических изображений, оказывается нерезким и с окрашенными краями. Это явление называется *хроматической аберрацией*, или *хроматизмом.*

Хроматическая аберрация устраняется путем комбинации линз, изготовленных из стекла с различными оптическими свойствами. Устранение ее для всех лучей спектра невозможно. Обычно совмещают изображения только для каких-либо двух лучей с различными длинами волн. Тогда говорят, что оптическая система *ахроматизована* или *исправлена в хроматическом отношении.* Оставшаяся хроматическая аберрация называется *вторичным спектром.* В большинстве случаев на практике она уже не сказывается существенно на качестве изображения. Выбор лучей, для которых должна быть ахроматизована оптическая система, определяется ее назначением. В случае визуальных приборов эти лучи надо выбирать по разные стороны вблизи желто-зеленой области спектра, к которой наиболее чувствителен человеческий глаз. Обычно в таких приборах ахроматизация производится для фраунгоферовых линий с длинами волн $\lambda_C = 656{,}3$ нм и $\lambda_F = 486{,}1$ нм. В фотографических аппаратах ахроматизация осуществляется для лучей, лежащих ближе к синему концу спектра, так как эти лучи сильнее действуют на фотографическую пластинку. Здесь обычно ахроматизация выполняется для фраунгоферовых линий с длинами волн $\lambda_D = 589{,}3$ нм и $\lambda_G = 434{,}0$ нм.

Для полной ахроматизации, т.е. совмещения изображений двух цветов (например, красных и синих), необходимо, чтобы были равны не только фокусные расстояния для этих цветов, но и совпадали соответствующие им главные плоскости. Однако во многих случаях достаточна уже *частичная ахроматизация*, т.е. либо равенство только фокусных расстояний без точного совмещения главных плоскостей, либо совмещение только главных плоскостей без точного равенства фокусных расстояний. Это зависит от назначения прибора и определяется тем, чтó в его работе важнее — увеличение изображения или его местоположение. Так, в окулярах зрительных труб и микроскопов главный интерес представляют углы, под которыми глаз видит различные окрашенные изображения предмета. Для равенства этих углов необходимо, чтобы окуляры были ахроматизованы в смысле одинаковости фокусных расстояний (см. § 24, п. 4).

Подробное рассмотрение ахроматизации дается в руководствах по расчету оптических систем. Здесь мы можем остановиться только на простейших принципиальных вопросах и притом только *в параксиальном приближении*. В высших приближениях также возникает хроматическая аберрация, устранение которой требует выполнения особых условий. Но мы не будем их рассматривать.

2. Ахроматизацию фокусного расстояния можно получить с помощью уже одной толстой линзы (см. задачу 2). Однако этот способ не имеет практического значения. Практически более важной является монохроматизация тонкой линзы или системы тонких линз.

Рассмотрим сначала одиночную тонкую линзу или систему тонких линз и подсчитаем, как меняется ее фокусное расстояние при малых изменениях показателя преломления. Взяв логарифмическую производную от выражения (10.9), получим

$$\frac{\delta f}{f} = -\frac{\delta n}{n-1}. \tag{16.1}$$

Так как показатель преломления уменьшается с увеличением длины волны, то из этой формулы следует, что фокусное расстояние f по абсолютной величине больше для красных лучей, чем для синих.

По сравнению с изменением фокусного расстояния положения главных плоскостей линзы меняются ничтожно. Рассмотрим, например, это изменение для главной плоскости пространства предметов. Для не слишком толстой линзы можно воспользоваться формулой (12.21), из которой находим: $\delta h/h = -\delta n/n$, и следовательно,

$$\frac{\delta h}{\delta f} = \frac{h}{f}\frac{n-1}{n} = -\frac{d}{R_2}\left(\frac{n-1}{n}\right)^2.$$

Так как даже для сравнительно толстой линзы, как правило, $d \ll R_2$ и, кроме того, $n-1 < n$, то $\delta h \ll \delta f$. Поэтому тонкая линза будет практически полностью ахроматизована, когда ахроматизовано ее фокусное расстояние.

В практических расчетах принимается, что формула (16.1) приближенно справедлива и для конечных изменений показателя преломления. В случае визуальных приборов полагают $\delta n =$

$= n_C - n_F$ и вводят величину

$$\nu = \frac{n_D - 1}{n_F - n_C}, \tag{16.2}$$

где n_D — показатель преломления для желтой D-линии натрия λ = 589,3 нм. Эта величина называется *коэффициентом дисперсии*, или *числом Аббе*. (Вместо n_D в числителе можно взять показатель преломления для любой другой длины волны из видимой области спектра, так как от этого ν изменится самое большее на 2 %) В результате формула (16.1) преобразуется к виду

$$\frac{\delta f}{f} = \frac{1}{\nu}, \tag{16.3}$$

откуда

$$\delta \frac{1}{f} = -\frac{1}{f\nu}. \tag{16.4}$$

В табл. 2 приведены оптические характеристики некоторых марок оптического стекла. Разность показателей преломления

Таблица 2

Марка стекла	n_D	$n_F - n_C$	$n_F - n_D$	$n_G - n_F$	ν
ЛК 6	1,4704	0,00704	0,00493	0,00339	66,8
К 1	1,4982	0,00765	0,00535	0,00423	65,1
К 3	1,5100	0,00805	0,00565	0,00451	63,4
К 14	1,5147	0,00849	0,00597	0,00477	60,6
БК 4	1,5302	0,00877	0,00617	0,00495	60,5
БК 6	1,5399	0,00905	0,00637	0,00513	59,7
ТК 2	1,5724	0,00996	0,00702	0,00567	57,5
ТК 21	1,6568	0,01285	0,00910	0,00745	51,1
ЛФ 1	1,5406	0,01145	0,00812	0,00673	47,2
ЛФ 7	1,5783	0,01407	0,01002	0,00843	41,1
Ф 6	1,6031	0,01590	0,01135	0,00903	37,9
Ф 8	1,6248	0,01757	0,1260	0,01088	35,6
БФ 1	1,5247	0,00955	0,00674	0,00547	54,9
БФ 28	1,6641	0,01874	0,01342	0,01145	35,4
ТФ 1	1,6475	0,01912	0,01369	0,01172	33,9
ТФ 5	1,7550	0,02743	0,01972	0,01728	27,5

Обозначения: ЛК — легкий крон, К — крон, БК — баритовый крон, ТК — тяжелый крон, ЛФ — легкий флинт, Ф — флинт, БФ — баритовый флинт, ТФ — тяжелый флинт.

$n_F - n_C$ называется *средней дисперсией*, отношения

$$(n_F - n_D)/(n_F - n_C), \qquad (n_G - n_F)/(n_F - n_C)$$

и аналогичные им — *относительными* или *частными дисперсиями*.

3. Рассмотрим теперь ахроматизацию тонкой линзы, сложенной из двух тонких линз, прижатых вплотную друг к другу. Обозначим через n_1 и f_1 показатель преломления и фокусное расстояние первой линзы, а через n_2 и f_2 — второй. Фокусное расстояние составной линзы находится по формуле

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}. \tag{16.5}$$

Оно будет одно и то же для длин волн λ_F и λ_C, если выполнено условие

$$\delta\frac{1}{f} = -\left(\frac{1}{f_1\nu_1} + \frac{1}{f_2\nu_2}\right) = 0, \tag{16.6}$$

или

$$f_1\nu_1 + f_2\nu_2 = 0. \tag{16.7}$$

Это и есть *условие ахроматизации составной линзы*. Так как коэффициенты дисперсии для всех оптических стекол положительны, то из равенства (16.6) следует, что фокусные расстояния f_1 и f_2 должны иметь *противоположные знаки*. Из уравнений (16.5) и (16.6) находим

$$\frac{1}{f_1} = \frac{1}{f}\frac{\nu_1}{\nu_1 - \nu_2}, \qquad \frac{1}{f_2} = -\frac{1}{f}\frac{\nu_2}{\nu_1 - \nu_2}. \tag{16.8}$$

Если общее фокусное расстояние f положительно, то линза с бо́льшим коэффициентом дисперсии будет иметь положительное, а линза с меньшим коэффициентом дисперсии — отрицательное фокусное расстояние. В частности, собирательная линза должна делаться из крона, а рассеивающая — из флинта (см. табл. 2).

Если задано фокусное расстояние f составной линзы, а также указаны сорта стекол, из которых изготовлены первая и вторая линзы, то фокусные расстояния f_1 и f_2 определятся из уравнений (16.8). Тем самым по формуле (10.9) определятся разности $\frac{1}{R_1^{(1)}} - \frac{1}{R_2^{(1)}}$ и $\frac{1}{R_1^{(2)}} - \frac{1}{R_2^{(2)}}$, где $R_1^{(1)}$ и $R_2^{(1)}$ — радиусы сферических поверхностей первой, а $R_1^{(2)}$ и $R_2^{(2)}$ — второй линз. Два из этих радиусов можно выбрать произвольно. Если линзы 1 и 2 должны склеиваться, то надо положить $R_2^{(1)} = R_1^{(2)}$. Остается еще один радиус (или, в общем случае, один свободный параметр). Им целесообразно распорядиться так, чтобы сделать сферическую аберрацию возможно меньше (см. задачу 1).

Ахроматические пары тонких склеенных линз широко применяются в объективах микроскопов. Каждая пара состоит из двояковыпуклой линзы из крона, склеенной с плоско-вогнутой линзой из флинта, плоская поверхность которой обращена в сторону падающего света.

4. Если составные части (подсистемы) сложной оптической системы *полностью ахроматизованы*, то будет *полностью ахроматизована* и система в целом. Действительно, так как первая подсистема полностью ахроматизована, то изображения, даваемые ею, совмещаются в двух цветах. Рассматривая совмещенные изображения как предмет для второй подсистемы, найдем, что изображения, даваемые и этой подсистемой, также совмещаются для тех же двух цветов. Применив это рассуждение для всех подсистем, найдем, что совмещение изображений будет иметь место и для системы в целом.

Обратное утверждение *не справедливо*. Система может быть полностью ахроматизована без того, чтобы были полностью ахроматизованы ее составные части. Однако если система состоит из двух тонких линз, находящихся на некотором расстоянии друг от друга, то справедливо и обратное утверждение. Действительно, пусть x_1, x'_1 и y_1, y'_1 — абсциссы и ординаты точки-предмета и ее промежуточного изображения, даваемого первой линзой, причем за начало координат принят центр этой линзы. Пусть x_2, y_2 и x'_2, y'_2 имеют тот же смысл для второй линзы, причем роль предмета в этом случае играет промежуточное изображение. Очевидно, ординаты промежуточного изображения y'_1 и y_2 одинаковы, а абсциссы x'_1 и x_2 связаны соотношением $x'_1 = x_2 + l$, где l — расстояние между линзами. Поперечные увеличения первой и второй линз равны соответственно

$$\frac{y'_1}{y_1} = \frac{x'_1}{x_1}, \qquad \frac{y'_2}{y_2} = \frac{x'_2}{x_2}.$$

Отсюда с учетом соотношения $y'_1 = y_2$ находим

$$\frac{y'_2}{y_1} = \frac{x'_1}{x_1}\frac{x'_2}{x_2}, \tag{16.9}$$

т.е. поперечное увеличение всей системы. По условию сложная система полностью ахроматизована, а потому увеличение должно быть одно и то же для обоих рассматриваемых нами цветов. Значит, вариация левой, а с ней и правой части равенства (16.9) обращается в нуль. Но абсцисса предмета x_1 постоянна, а абсцисса окончательного изображения x'_2 одинакова для обоих цветов. Поэтому, взяв вариацию логарифма от правой части (16.9), получим: $\frac{\delta x'_1}{x'_1} - \frac{\delta x_2}{x_2} = 0$. Кроме того, из соотношения $x'_1 = x_2 + l$, ввиду постоянства расстояния l между линзами, следует: $\delta x_1' = \delta x_2$. Таким образом, $\left(\frac{1}{x_2'} - \frac{1}{x_2}\right)\delta x_1' = 0$. Если $\frac{1}{x_1'} - \frac{1}{x_2} \neq 0$, т.е. $l \neq 0$, то $\delta x_1' = 0$, т.е. промежуточные изображения совмещаются в обоих цветах. Но это означает, что первая линза полностью ахроматизована. Учтя свойство обратимости световых лучей, убедимся, что полностью ахроматизована и вторая линза. Исключение составляет случай, когда $\frac{1}{x_1'} - \frac{1}{x_2} = 0$. В этом случае линзы вплотную прижаты друг к другу ($l = 0$), и ахроматизация возможна даже тогда, когда каждая из линз не ахроматизована. Такой случай был разобран выше.

5. Таким образом, *чтобы система из двух тонких линз, расстояние l между которыми не равно нулю, была полностью ахроматизована, необходимо, чтобы были в отдельности полностью ахроматизованы обе составляющие ее линзы.* Однако *если не требовать полной ахроматизации системы, а ограничиться только ахроматизацией ее фокусных расстояний, то такая ахроматизация может быть получена и с неахроматизованными линзами.*

Действительно, фокусное расстояние системы из двух тонких линз определяется формулой (12.8). Из нее находим условие ахроматизации фокусного расстояния:

$$\delta\frac{1}{f} = \left(\frac{1}{f_1} - \frac{l}{f_1 f_2}\right)\frac{\delta n_1}{n_1 - 1} + \left(\frac{1}{f_2} - \frac{l}{f_1 f_2}\right)\frac{\delta n_2}{n_2 - 1} = 0, \qquad (16.10)$$

где n_1 — показатель преломления первой линзы, а n_2 — второй. Если n_1 и n_2 одинаковы, то это условие переходит в

$$l = \frac{1}{2}(f_1 + f_2). \qquad (16.11)$$

Таким образом, система из двух тонких линз, изготовленных из одного и того же стекла, будет ахроматизована в отношении фокусного расстояния и притом для всех цветов спектра, если расстояние между линзами равно полусумме их фокусных расстояний. Такой способ ахроматизации применяется в окулярах зрительных труб (см. § 24).

6. Объективы, в которых ахроматизация выполнена для двух цветов спектра, называются *ахроматами*. В некоторых оптических системах (в объективах микроскопов, длиннофокусных астрономических труб и спектральных аппаратов) наличие вторичного спектра (см. п. 1) существенно ухудшает качество изображения. Аббе ввел в микроскопию *апохроматы*, т.е. объективы, в которых ахроматизация выполнена для трех цветов спектра и, кроме того, удовлетворено условие синусов (см. § 18). Остающаяся при этом хроматическая аберрация называется *третичным спектром*.

7. В заключение упомянем о *сложных призмах*, изготовляемых из стекол различных сортов. На рис. 60 изображена ахроматическая призма. Первая призма, из крона, отклоняет и рассеивает лучи. Вторая призма, из флинта, отклоняет их в противоположную сторону. Рассеивающая способность (дисперсия) флинта больше, чем крона. Поэтому вторую призму можно подобрать так, чтобы компенсировать рассеивающую способность первой призмы, т.е. добиться одинаковости направлений какихлибо двух лучей (например, красного и фиолетового) при их выходе из рассматриваемой системы. Однако отклонение лучей, вызываемое первой призмой, компенсируется лишь частично. В результате белый луч проходит через сложную призму без заметной дисперсии, но испытывает отклонение в сторону.

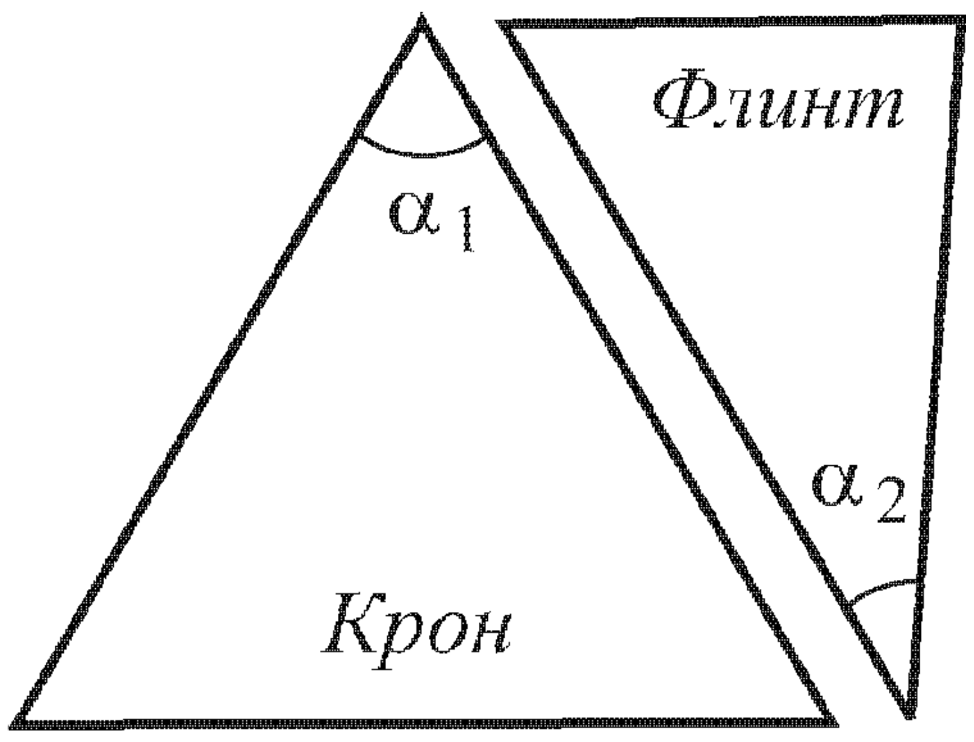
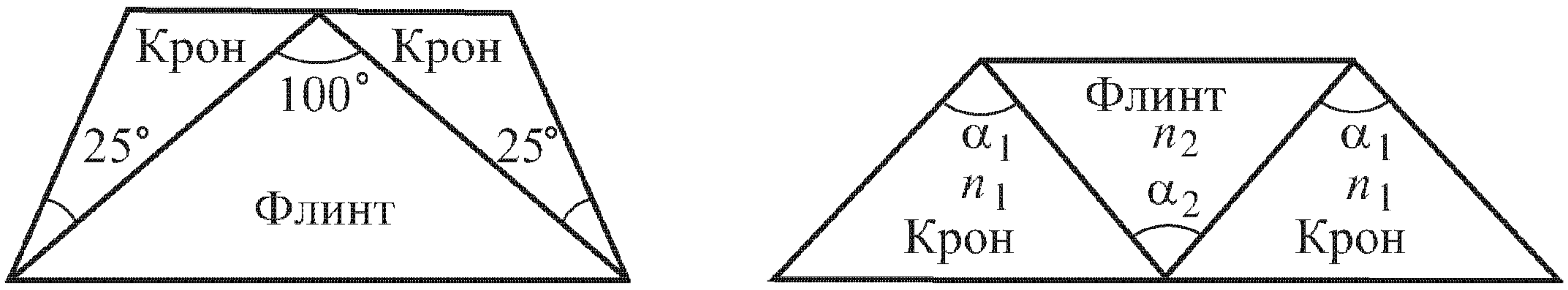


Рис. 60

На рис. 61 *а* изображена *дисперсионная призма*, применяемая в спектрографах. У нее дисперсия очень значительна благодаря большому преломляющему углу внутренней призмы из флинта. Боковые призмы из крона мало влияют на общую дисперсию призмы, но сильно уменьшают отклонение лучей, так как угол между их наружными гранями сравнительно невелик. Кроме того, эти призмы позволяют увеличить преломляющий угол внутренней призмы, величина которого лимитируется полным отражением. Наконец, на рис. 61 *б* изображена *призма прямого зрения*. Она производит спектральное разложение, но оставляет неотклоненным средний луч пучка света.



а *б*

Рис. 61

ЗАДАЧИ

1. Рассчитать ахроматический объектив с фокусным расстоянием $f = 50$ см, склеенный из двух линз. Передняя линза изготовлена из крона К 1, задняя — из флинта Ф 8. Задняя линза — выпукло-вогнутая, обращена выпуклой поверхностью наружу, радиус кривизны этой поверхности равен 100 см.

Решение. Из табл. 2 находим для первой линзы: $\nu_1 = 65{,}1$, $n_D = 1{,}4982$; для второй линзы $\nu_2 = 35{,}6$, $n_D = 1{,}6248$. По формулам (16.8) вычисляем фокусные расстояния линз:

$$f_1 = f\frac{\nu_1 - \nu_2}{\nu_1} = 50 \cdot \frac{65{,}1 - 35{,}6}{65{,}1} = 22{,}66 \text{ см},$$

$$f_2 = -f\frac{\nu_1 - \nu_2}{\nu_2} = -50 \cdot \frac{65{,}1 - 35{,}6}{35{,}6} = -41{,}43 \text{ см}.$$

По условию задачи $R_2^{(2)} = -100$ см. По формуле (10.9) находим $R_1^{(2)}$:

$$R_1^{(2)} = \frac{(n-1)f_2R_2^{(2)}}{(n-1)f_2 + R_2^{(2)}} = \frac{0{,}6248 \cdot 41{,}43 \cdot 100}{0{,}6248 \cdot 41{,}43 - 100} = -34{,}93 \text{ см}.$$

Той же величине равен радиус $R_2^{(1)}$, так что

$$R_1^{(1)} = \frac{(n-1)f_1R_2^{(1)}}{(n-1)f_1 + R_2^{(1)}} = -\frac{0{,}4982 \cdot 22{,}66 \cdot 20{,}56}{0{,}4982 \cdot 22{,}66 - 20{,}56} = 25{,}04 \text{ см}.$$

Вид рассчитанного объектива изображен на рис. 62.

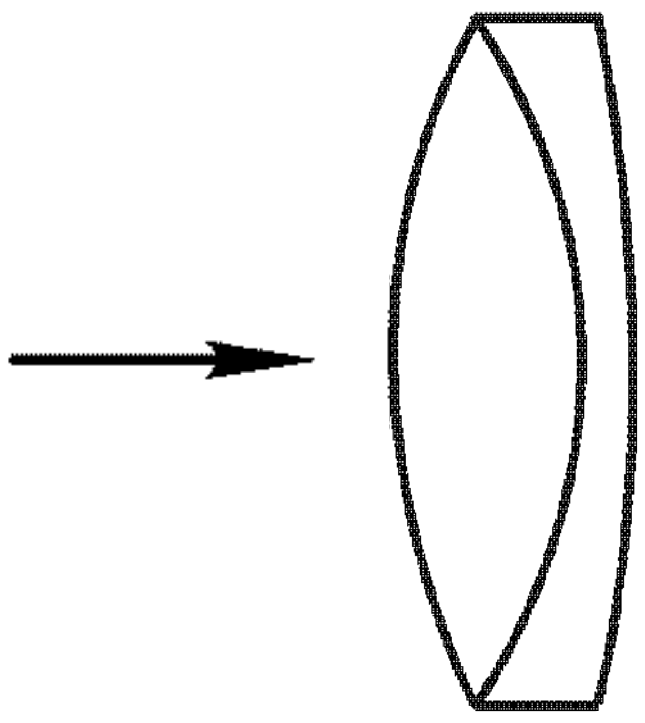


Рис. 62

2. Возможна ли ахроматизация толстой одиночной линзы относительно фокусных расстояний для двух длин волн?

Ответ. Ахроматизация возможна потому, что в формулу для фокусного расстояния толстой линзы входит n^2. Следовательно, для заданного фокусного расстояния могут существовать два значения показателя преломления n_1 и n_2, которым соответствует это фокусное расстояние. Толщина линзы должна быть равна

$$d = \frac{n^2}{n^2 - 1}(R_1 - R_2),$$

где $n = \sqrt{n_1n_2}$ — среднее геометрическое из n_1 и n_2. Так как толщина d существенно положительна, то для возможности ахроматизации необходимо, чтобы $R_1 - R_2 > 0$. Этому условию удовлетворяют только двояковыпуклые и плоско-вогнутые линзы. Все прочие толстые линзы не могут быть ахроматизованы.

3. Написать условие ахроматизации преломляющей призмы для двух близких цветов. Призма состоит из двух призм с малыми преломляющими углами, изготовленных из стекол с коэффициентами дисперсии ν_1 и ν_2.

Ответ. Призмы должны быть обращены преломляющими углами в противоположные стороны. Условие ахроматизации: $\frac{\varphi_1}{\nu_1} - \frac{\varphi_2}{\nu_2} = 0$, где φ_1 — угол отклонения луча, даваемый первой призмой, а φ_2 — второй. Если $\nu_1 \neq \nu_2$, то результирующее отклонение $\varphi_1 - \varphi_2$ отлично от нуля.

§ 17. Условие отсутствия дисторсии

1. В фотографическом объективе особо важное значение имеет устранение изгиба плоскости изображения и дисторсии. Найдем *условие отсутствия дисторсии*. Пусть PP_1P_2

(рис. 63) — плоскость предмета, $P'P_1'P_2'$ — сопряженная ей плоскость изображения, O и O' — центры входного и выходного зрачков. Проведем главные лучи P_1O и P_2O от точек предметной плоскости через центр входного зрачка. Сопряженные им лучи $P_1'O'$ и $P_2'O'$ пройдут через центр O' выходного зрачка. Обозначим через u и u' углы наклона этих лучей к главной оптической оси. Тогда $PP_1 = PO\operatorname{tg}u_1$, $P'P_1' = P'O'\operatorname{tg}u_1'$, $PP_2 = PO\operatorname{tg}u_2$, $P'P_2' = P'O'\operatorname{tg}u_2'$, и следовательно,

$$\frac{P'P_1'}{PP_1} = \frac{\operatorname{tg}u_1'}{\operatorname{tg}u_1}, \quad \frac{P'P_2'}{PP_2} = \frac{\operatorname{tg}u_2'}{\operatorname{tg}u_2}.$$

Отсутствие дисторсии означает, что поперечные увеличения $P'P_1'/PP_1$ и $P'P_2'/PP_2$ одинаковы при любых положениях точек P_1 и P_2. Следовательно, для устранения дисторсии необходимо и достаточно, чтобы соблюдалось равенство

$$\frac{\operatorname{tg}u_1'}{\operatorname{tg}u_1} = \frac{\operatorname{tg}u_2'}{\operatorname{tg}u_2}, \tag{17.1}$$

каковы бы ни были значения углов u_1 и u_2. Это равенство называется *условием тангенсов*, или *условием ортоскопии*. Точки O и O' главной оптической оси, удовлетворяющие этому условию, называются *ортоскопическими точками*. Центры входного и выходного зрачков являются центрами перспективы предмета и его изображения. Таким образом, условие отсутствия дисторсии сводится к требованию, чтобы эти центры перспективы были ортоскопическими точками.

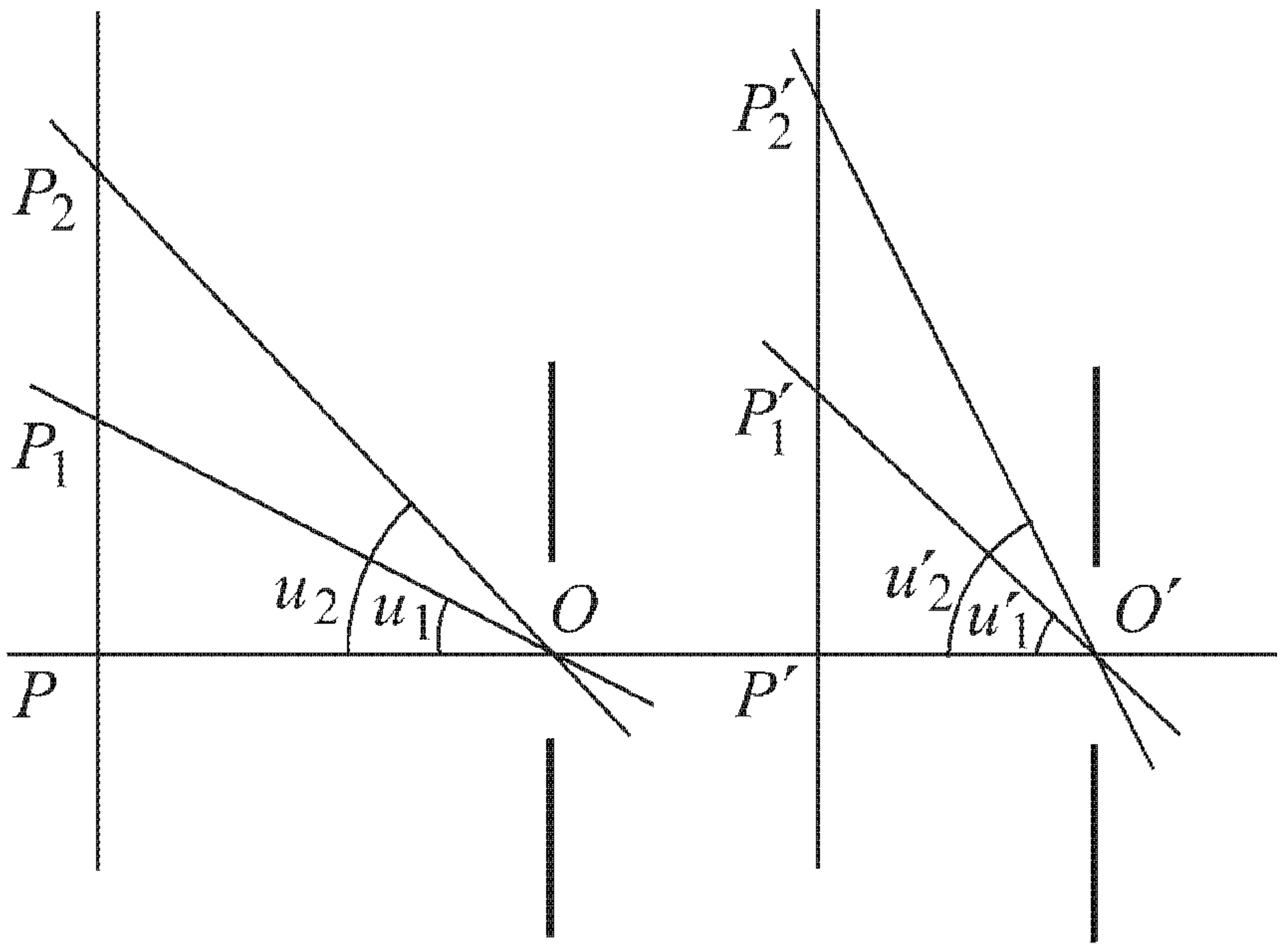


Рис. 63

2. Простейшим примером, где выполнено условие ортоскопии, может служить симметричный фотографический объектив, изображенный на рис. 64. Он состоит из двух совершенно одинаковых ахроматизованных двойных линз, обращенных друг к другу вогнутыми поверхностями. Апертурная диафрагма помещена посередине между этими линзами. Так как она расположена близко к обеим линзам, то оба изображения ее мнимые, прямые и лежат обычно внутри объектива недалеко от диафрагмы A_1A_2. Изображение D_1D_2, даваемое передней линзой, лежит справа от диафрагмы A_1A_2 и служит входным зрачком системы; изображение $D_1'D_2'$, даваемое задней линзой, лежит левее диафрагмы A_1A_2 и служит выходным зрачком. Выходной зрачок

$D'_1D'_2$ есть изображение входного зрачка D_1D_2, даваемое всей оптической системой.

Как видно из рис. 64, это изображение производится с поперечным увеличением +1. Поэтому плоскости входного и выходного зрачков являются главными плоскостями, а центры этих зрачков — главными точками объектива. Возьмем теперь любой падающий луч, продолжение которого проходит через центр входного зрачка. Тогда, ввиду симметрии, продолжение выходящего луча пройдет через центр выходного зрачка, а самый луч внутри системы — через центр апертурной диафрагмы. Следовательно, углы наклона u и u' этих главных лучей к главной оптической оси будут всегда одинаковы, каковы бы ни были значения самих углов u и u'. Это значит, что в рассматриваемой системе соблюдено условие ортоскопии.

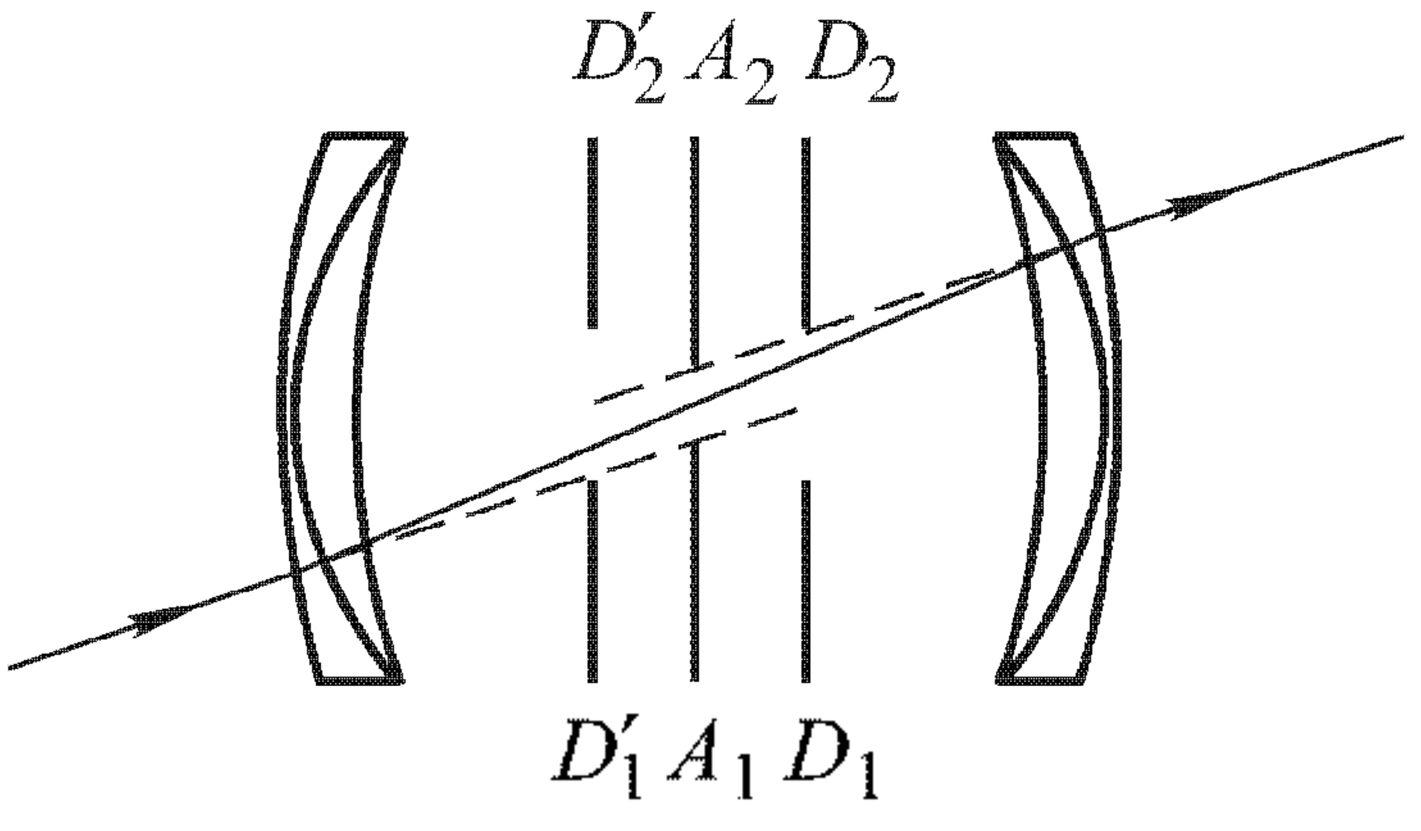


Рис. 64

§ 18. Условие синусов Аббе

1. В микроскопах изучаемые объекты малы и помещаются перед объективом вблизи главной оптической оси. Апертуры лучей, формирующих изображение в микроскопе, должны быть как можно шире (в лучших объективах микроскопов апертуры практически доведены до своего теоретического предела 180°). От этого увеличивается яркость изображения, а главное повышается разрешающая сила объектива, т.е. его способность различать мелкие детали (см. § 56). Поэтому сопряженные точки главной оптической оси, в первой из которых помещается объект, а во второй получается его изображение, даваемое объективом, должны быть *анаберрационными*, т.е. точка P должна изображаться в виде точки P' широкими пучками лучей (см. § 9). Однако этого условия еще недостаточно. Необходимо, чтобы без аберраций изображались все точки малого участка предметной плоскости, проходящей через точку P перпендикулярно к главной оптической оси. Практически речь идет об устранении сферической аберрации и комы. Анаберрационные точки главной оптической оси P и P', удовлетворяющие этому условию, называются *апланатическими*.

Значение апланатизма было выяснено немецким физиком Аббе (1840–1905). Аббе был оптиком-теоретиком. Он не только глубоко понимал теорию оптических приборов, но и указал пути их совершенствования. Особое внимание он уделил теории и совершенствованию микроскопа. Его многолетнее плодотворное сотрудничество с известной оптической фирмой Цейсса (1816–1888) в Иене (Германия) привело к созданию и массовому производству лучших в мире оптических приборов, в которых практически было достигнуто то, что принципиально возможно с точки зрения волновой теории света. Аббе вывел необходимое и достаточное условие апланатизма, получившее название *условия синусов*. К выводу этого условия мы и перейдем.

2. Ввиду осевой симметрии оптической системы, при выводе достаточно ограничиться рассмотрением изображений не плоских площадок, а *прямолинейных отрезков* малой, но произвольной длины, перпендикулярных к главной оптической оси. Пусть один из таких отрезков, PQ, изображается широкими пучками лучей в виде отрезка $P'Q'$, также перпендикулярного к главной оптической оси (рис. 65). Оптические длины всех лучей, соединяющих сопряженные точки P и P', одинаковы. То же справедливо для лучей, соединяющих сопряженные точки Q и Q'.

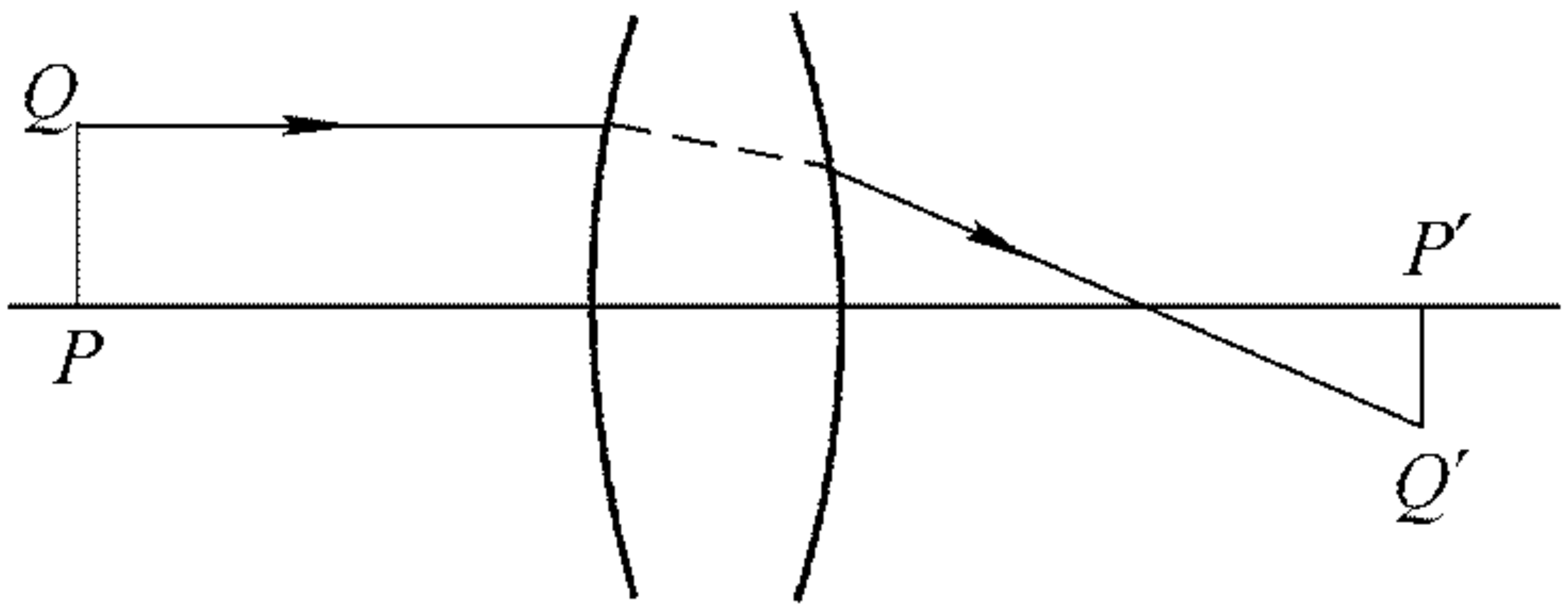


Рис. 65

Докажем, что оптическая длина (PP') любого луча, соединяющего апланатические точки P и P', равна оптической длине (QQ') любого луча, соединяющего сопряженные точки Q и Q'. Иными словами, оптические длины лучей между сопряженными точками отрезка PQ и его изображения $P'Q'$ одинаковы для всех пар сопряженных точек. Ввиду отмеченного выше свойства оптических длин лучей, соединяющих сопряженные точки, достаточно доказать это утверждение для параксиальных лучей, выходящих из точек объекта PQ параллельно главной оптической оси. Волновой фронт, соответствующий таким лучам, после прохождения через объектив становится сферическим, малый участок которого можно считать плоским. Таким образом, объект PQ целиком лежит в плоскости падающего, а его изображение $P'Q'$ — в плоскости прошедшего волновых фронтов. Но оптические длины всех лучей между двумя положениями волнового фронта одинаковы. Отсюда и следует доказываемое утверждение.

Проведем теперь через точки отрезка PQ пучок параллельных лучей под произвольным углом u к главной оптической оси (рис. 66). Так как лучи лежат в одной плоскости, то соответствующий им волновой фронт представится прямолинейным отрезком QA, перпендикулярным к лучам. После прохождения через объектив параллельность лучей нарушится. Однако ввиду малости объекта PQ будет мал и угол расхождения прошедших лучей. Пренебрегая им, можно считать, что все прошедшие лучи наклонены к оптической оси практически под одним и тем же углом u'. Соответствующий участок волнового фронта представится прямолинейным отрезком $Q'A'$. Ввиду одинаковости оптических длин всех лучей между волновыми фронтами $(AP'A') = (QQ')$. По доказанной только что теореме $(PAP') = (QQ')$, и предыдущее соотношение переходит в $(PAP') = (AP'A')$. Вычитая из обеих частей этого равенства общую часть (AP'), получим: $(PA) = (P'A')$, или

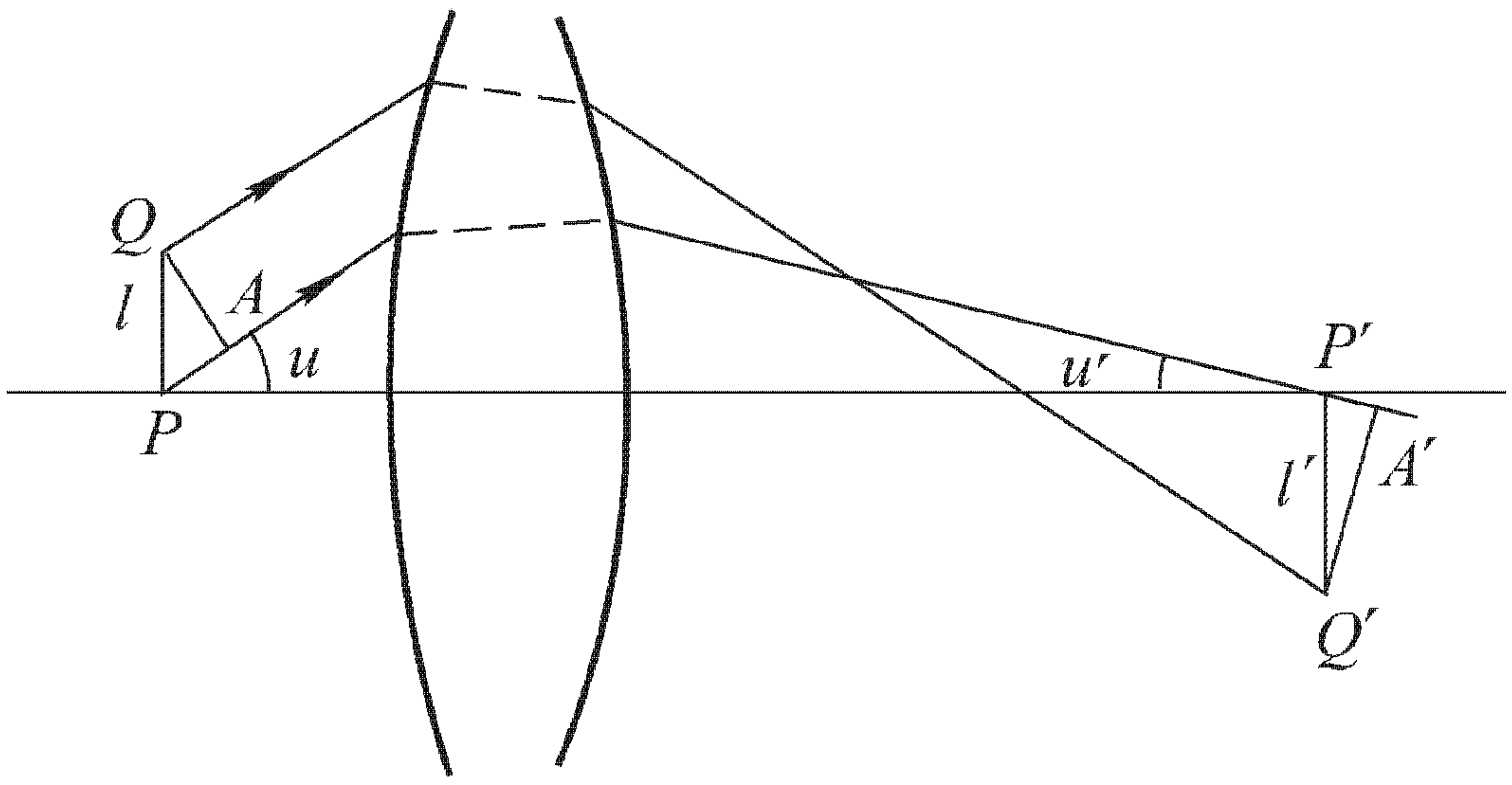


Рис. 66

$$nl \sin u = n'l' \sin u', \tag{18.1}$$

где l и l' — длины отрезков PQ и $P'Q'$, а n и n' — показатели преломления пространств предметов и изображений. Равенство (18.1) и выражает условие синусов Аббе. Оно должно выполняться для любых значений угла u и для любых малых значений длины отрезка l. Для параксиальных лучей условие синусов (18.1) переходит в теорему Лагранжа–Гельмгольца (10.6) и, следовательно, выполняется для любой пары сопряженных точек на главной оптической оси.

В следующем параграфе приводится более строгий вывод условия синусов, основанный на теореме косинусов.

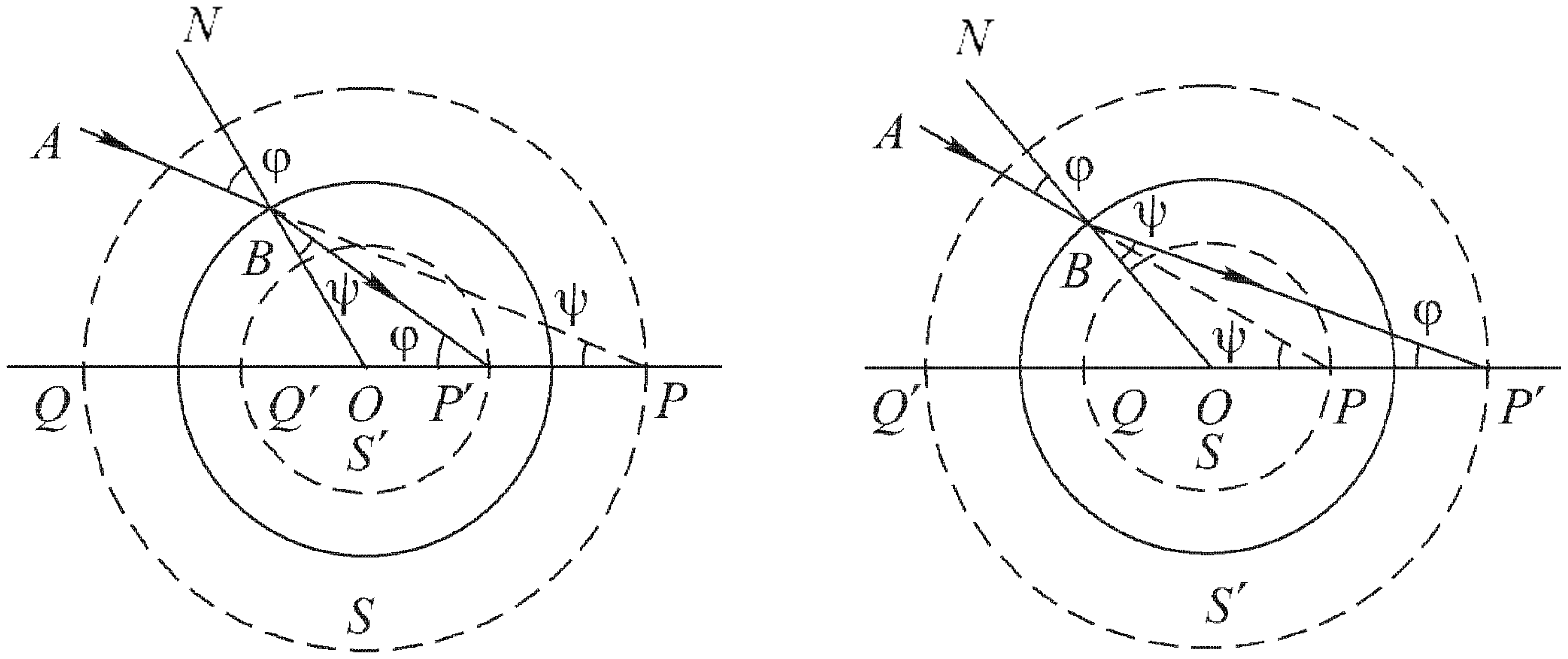
3. К важному примеру апланатических точек приводит построение преломленного луча на поверхности шара, указанное Вейерштрассом (1815–1897). Пусть R — радиус шара, n — его показатель преломления относительно окружающей среды (рис. 67). Построим две концентрические с шаровой поверхностью сферы S и S' с радиусами nR и R/n. Продолжим падающий луч AB до пересечения в точке P со сферой S и соединим точку P с центром шара O. Прямая OP пересечет сферу S' в точке P'. Прямая, соединяющая точку падения B с точкой P', дает направление преломленного луча. Действительно, из подобия треугольников OBP и OBP' следует, что угол $OP'B$ равен углу падения φ, а потому

$$\frac{\sin\varphi}{\sin\psi} = \frac{OB}{OP'} = n,$$

что и доказывает наше утверждение.

Из построения видно, что падающий пучок лучей, сходящийся в точке P, после преломления на шаровой поверхности будет сходиться в точке P'. Обратно, если точечный источник поместить в P', то после преломления получится пучок лучей, расходящихся из P. Следовательно, P и P' являются сопряженными анаберрационными точками и притом апланатическими.

Действительно, ввиду шаровой симметрии не только точки P и P', но и сферы S и S' отображаются друг в друга широкими пучками лучей. Так как обе сферы нормальны к прямой PP', то точки P и P' должны удовлетворять условию синусов, в чем легко убедиться и непосредственно. Точки Q и Q', очевидно, также апланатические. Наконец, центр сферы O можно рассматривать как пару совпадающих апланатических точек, являющихся одновременно узловыми точками системы.



$n>1$
а

$n<1$
б

Рис. 67

Итак, на оптической оси QP имеется три пары изолированных апланатических точек: P и P', Q и Q' и двойная апланатическая точка O.

4. Апланатические точки шара используются для построения *иммерсионных объективов* микроскопов с очень большими увеличениями. Иммерсионным называют такой объектив, когда между покровным стеклом и фронтальной линзой объектива находится слой жидкости (*иммерсия*). Иммерсия была предложена Амичи (1786–1863) в 1840 г. С целью уменьшения вредного для качества изображения преломления на поверхности покровного стекла он заполнял водой промежуток между покровным стеклом и фронтальной линзой объектива. Аббе в 1878 г. стал применять *однородную иммерсию* с показателем преломления, почти равным показателю преломления покровного стекла и фронтальной линзы. В такой системе лучи, исходящие из каждой точки предмета, распространяются практически прямолинейно до выхода из фронтальной линзы объектива. В качестве однородной иммерсии применяется кедровое масло ($n = 1{,}515$). В однородной иммерсионной системе Аббе фронтальная линза объектива состоит из стеклянного полушария, плоская сторона которого обращена к предмету (рис. 68). Предмет помещается на расстоянии $OP = R/n$ от центра (R — радиус полушария). Так как лучи до выхода из фронтальной линзы не испытывают преломления, то P будет апланатической точкой. Изображение предмета получается в сопряженной апланатической точке P' на расстоянии $OP' = Rn$ от центра O. Как видно из построения Вейерштрасса, линейное увеличение равно n^2. Поэтому из (18.1) следует $\sin u' = (\sin u)/n$, так что угол наклона луча к оптической оси уменьшается.

Рис. 68

Амичи для получения большего увеличения и дальнейшего уменьшения углов наклона лучей к оптической оси предложил помещать за фронтальной линзой вогнуто-выпуклую линзу. Точка P' должна находиться в центре кривизны вогнутой поверхности линзы. По отношению к преломлению на этой поверхности P' будет апланатической точкой, совпадающей со своей сопряженной точкой. Точка P' должна одновременно находиться на расстоянии R_1/n_1 от центра кривизны выпуклой поверхности второй линзы (n_1 — показатель преломления этой линзы, R_1 — радиус кривизны ее выпуклой поверхности). Тогда по отношению к преломлению на этой поверхности P' будет апланатической точкой; ее изображение получится в сопряженной апланатической точке P''.

Применив метод Амичи несколько раз, можно добиться какого угодно уменьшения углов наклона лучей к оптической оси.

При этом для всей системы в целом будет выполнено условие синусов. Метод Амичи часто применяют при конструкции объективов микроскопов. Однако таким методом конструируются в лучшем случае первые две линзы, так как иначе возникает сильная, ничем не компенсируемая хроматическая аберрация.

5. Аббе принадлежит простой способ испытания объективов на выполнение условия синусов. Допустим, что во второй апланатической точке объектива P' (рис. 69) помещена малая диафрагма, через которую производится наблюдение. Ее сильно уменьшенное изображение, получающееся в сопряженной апланатической точке P, будет входным зрачком системы.

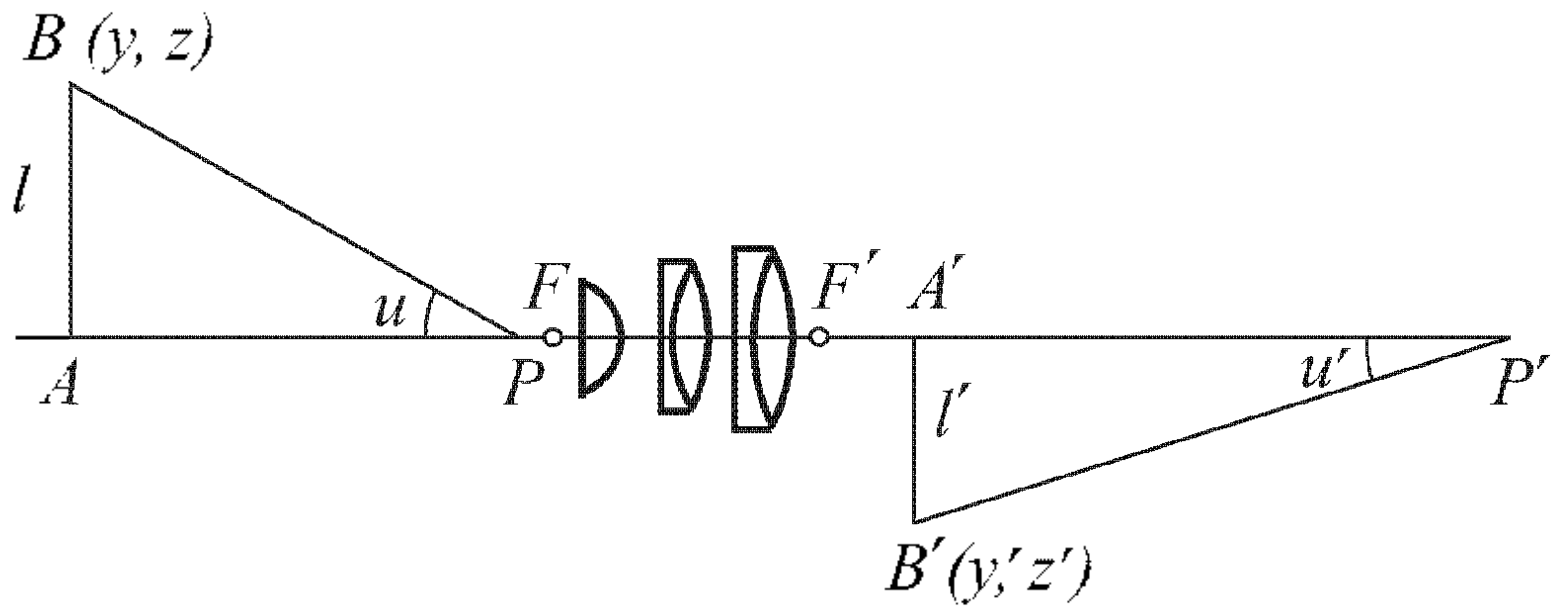


Рис. 69

Рассмотрим изображение конечного участка плоскости AB, даваемое объективом, когда расстояние q этой плоскости от точки P очень велико по сравнению с диаметром входного зрачка. Так как при перемещении объекта влево от апланатической точки P его изображение перемещается также влево (см. § 11, п. 6), то изображение плоскости AB получится левее второй апланатической точки P'. Когда расстояние q достаточно велико, то изображение $A'B'$ плоскости AB можно увидеть невооруженным глазом, если смотреть через диафрагму. Изображение получится сильно искаженным из-за использования косых пучков лучей, наклоненных под большими углами к главной оптической оси. Однако, ввиду исключительной малости входного зрачка, в разложении аберраций в степенные ряды можно пренебречь всеми степенями радиуса входного зрачка σ, сохранив лишь член нулевой степени, не содержащий σ. Иными словами, достаточно ограничиться учетом *одной только дисторсии*. В этом приближении изображение $A'B'$ будет расположено в плоскости параксиальных изображений.

Поставим задачу найти в предметной плоскости AB такое семейство кривых, чтобы его изображение в плоскости $A'B'$ получилось в форме квадратной сетки.

Возьмем в предметной плоскости произвольную точку B с координатами y, z на расстоянии $\sqrt{y^2+z^2}$ от главной оптической оси. Проведем луч BP через центр входного зрачка. Сопряженный луч $B'P'$ однозначно определится условием синусов $\sin u'/\sin u = A$, где $A = nl/(n'l')$, т.е. для рассматриваемого объектива A — величина постоянная, u и u' — углы наклона падающего и прошедшего лучей к главной оптической оси. Положение изображения B' определится точкой пересечения луча $B'P'$ с плоскостью параксиальных изображений. Подставим вместо $\sin u$ и $\sin u'$ их значения

$$\sin u = \sqrt{\frac{y^2+z^2}{q^2+y^2+z^2}}, \quad \sin u' = \sqrt{\frac{y'^2+z'^2}{q'^2+y'^2+z'^2}},$$

где q' — расстояние плоскости $A'B'$ от апланатической точки P'. Учтем при этом, что точки B и B' лежат в одной плоскости, проходящей через главную оптическую ось, так что $z'/y' = z/y$. Тогда получим

$$[A^2q'^2 - (1 - A^2)y'^2]y^2 - y'^2(1 - A^2)z^2 = q^2y'^2.$$

Возьмем теперь в плоскости $A'B'$ семейство прямых, параллельных оси Z: $y' = C$, где C — параметр, который может принимать произвольные значения. Уравнение кривых в предметной плоскости AB, изображением которых является это семейство, найдется из предыдущего соотношения, если в нем y' заменить на C. Таким путем получим

$$\frac{y^2}{a^2} - \frac{z^2}{b^2} = 1, \tag{18.2}$$

где

$$a^2 = \frac{q^2\omega^2}{1 - (1 - A^2)\omega^2}, \quad b^2 = \frac{q^2}{1 - A^2}, \tag{18.3}$$

а ω — новый параметр, связанный с параметром C соотношением

$$\omega = C/Aq'. \tag{18.4}$$

Для объективов микроскопов всегда $A^2 < 1$, так что при $\omega^2 < 1/(1-A^2)$ величина a существенно положительна. В этом случае (18.2) представляет семейство гипербол, осью которого является координатная ось Y. При $\omega = 0$ уравнение (18.2) переходит в $y^2 = 0$ и изображает ось Z. Значения $\omega^2 > 1/(1 - A^2)$ надо исключить, так как в этом случае уравнение (18.2) представляет семейство мнимых эллипсов и никакого оптического изображения не получается. Аналогично, семейство прямых $z' = C$ при тех же значениях параметра ω является оптическим изображением семейства гипербол

$$\frac{z^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{18.5}$$

получающегося из семейства (18.2) поворотом на 90° вокруг начала координат O. Оба семейства гипербол (18.2) и (18.5) пересекаются между собой, образуя криволинейную сетку, представленную на рис. 70. При переходе от одной гиперболы к соседней параметр ω должен получать одинаковые приращения. Тогда и приращения параметра C, как это видно из (18.4), также будут одинаковыми. Следовательно, если рис. 70 поместить на расстоянии q перед передней апланатической точкой P, то гиперболы изобразятся в виде квадратной сетки прямых. При $\omega^2 = (1 - A^2)/2$ уравнение (18.2) переходит в $y^2 - z^2 = b^2$, а уравнение (18.5) — в $z^2 - y^2 = b^2$.

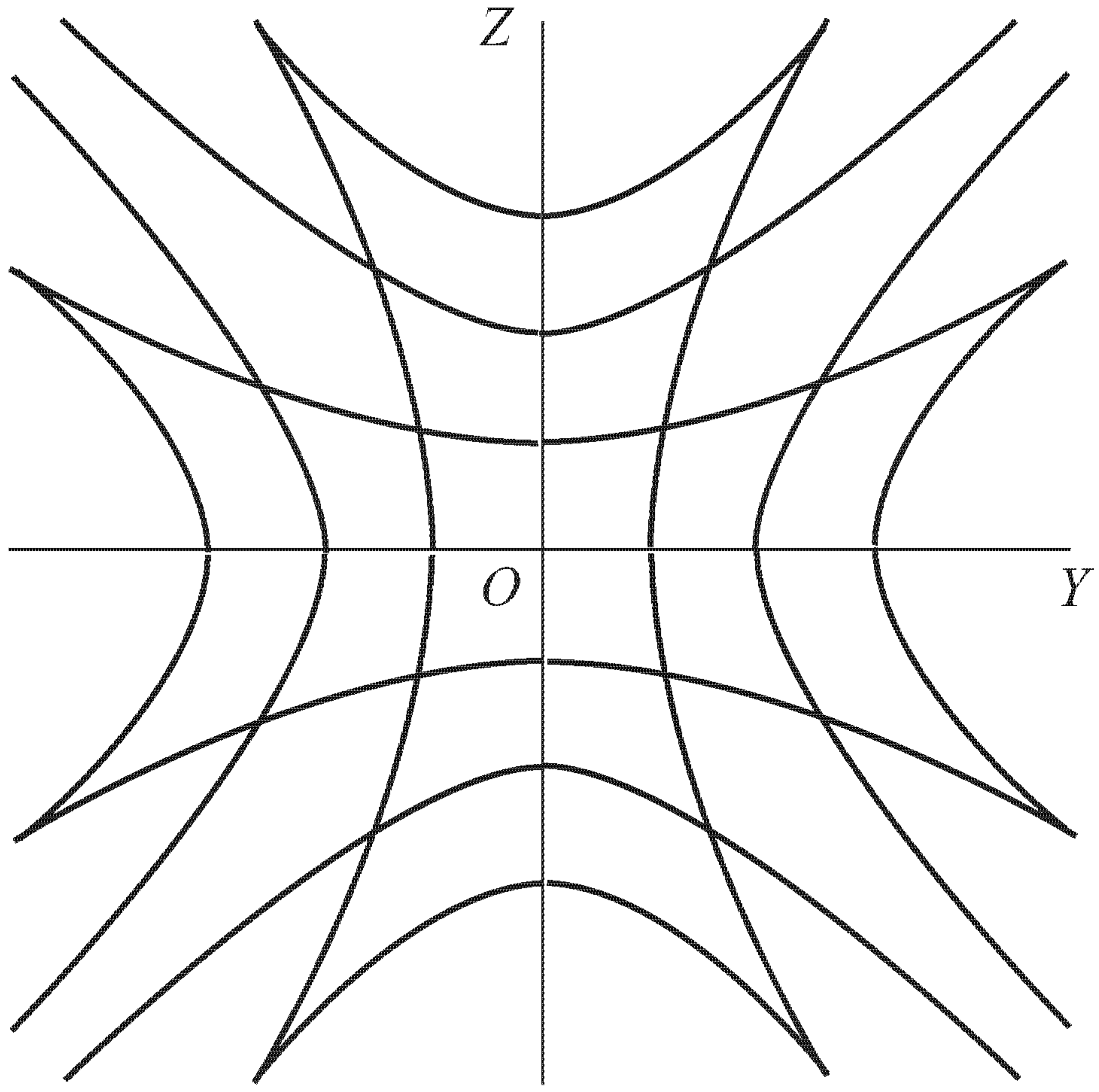


Рис. 70

Обе эти гиперболы имеют асимптотами биссектрисы координатных углов. Следовательно, b есть расстояние от начала координат до вершины той из гипербол, асимптоты которой совпадают с биссектрисами координатных углов. Для объективов микроскопов A — величина малая, квадратом которой по сравнению с единицей можно пренебречь. В этом приближении $b = q$. Это дает простой способ определения расстояния q, на котором следует поместить рис. 70, чтобы его изображение получилось в виде квадратной сетки.

Теперь становится понятным метод Аббе испытания объективов микроскопов на выполнение условия синусов. Аббе пользовался шаблоном, воспроизведенным на рис. 71. Удалив окуляр микроскопа, следует поместить такой шаблон на расстоянии q перед передней апланатической точкой. Глаз наблюдателя помещается во вторую апланатическую точку. Зрачок глаза играет роль выходного, а его изображение, даваемое объективом, — входного зрачка системы. Если изображение шаблона получается в виде сетки квадратов, то объектив удовлетворяет условию синусов. Если изображение получается слишком мелким, то можно применить вспомогательный микроскоп небольшого увеличения, перед объективом которого помещена малая диафрагма, расположенная во второй апланатической точке P'. При испытании описанным способом большого числа объективов микроскопов, ранее изготовленных лучшими мастерами без всяких расчетов путем последовательных проб и подбора линз, Аббе обнаружил, что все хорошие объективы всегда удовлетворяли условию синусов.

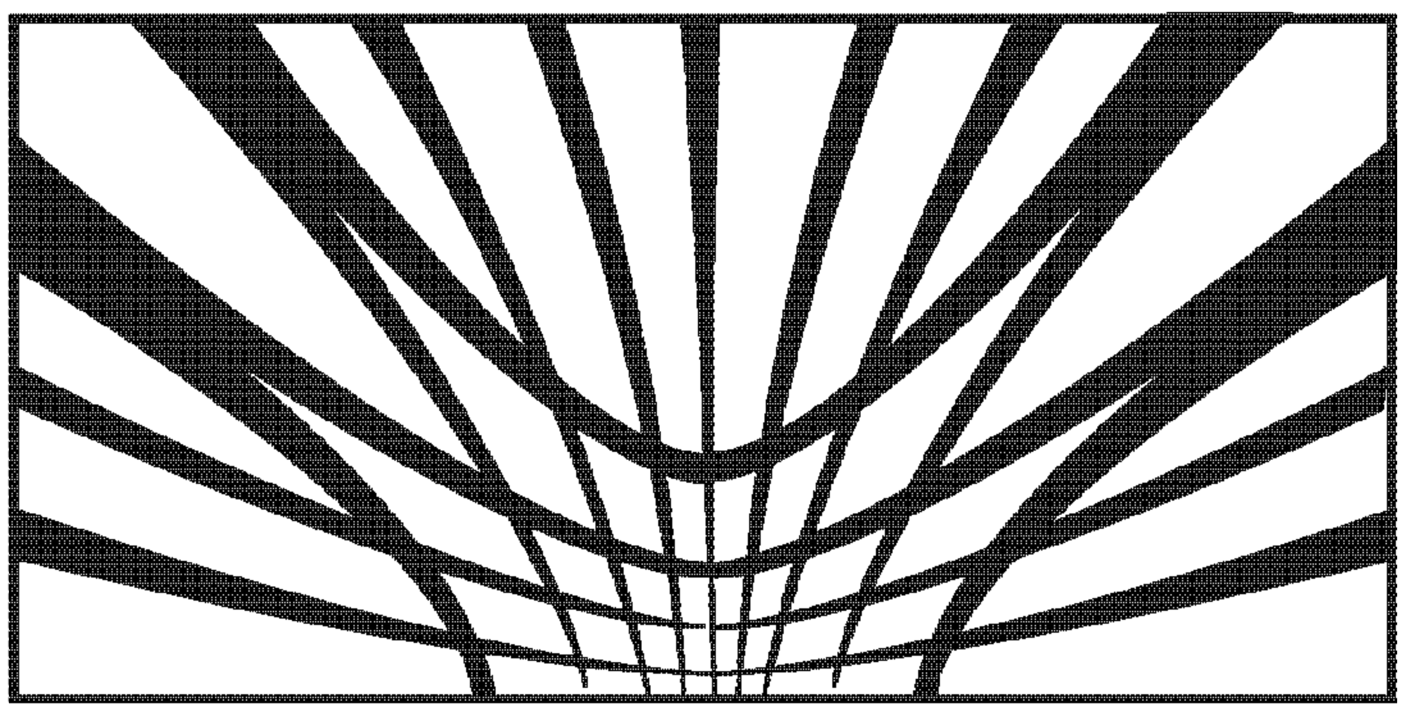


Рис. 71

§ 19. Теорема косинусов. Стигматические изображения широкими пучками лучей

1. Обобщим результаты предыдущего параграфа на случай произвольных оптических систем. Среды, в которых распространяются световые лучи, здесь предполагаются изотропными, но могут быть неоднородными. Таким образом, в общем случае световые лучи будут криволинейными. Пусть P и P' — две точки с радиусами-векторами $\mathbf{r}$ и $\mathbf{r}'$, лежащие на одном луче. Оптическая длина луча, соединяющего эти точки, рассматриваемая как функция их координат, называется *точечным эйконалом*, или *характеристической функцией оптической системы*. Она была введена Гамильтоном (1805–1865) и оказалась весьма полезной при исследовании оптических изображений. Характеристическую функцию будем обозначать через $H = H(\mathbf{r}, \mathbf{r}')$.

Фиксировав положение начальной точки P, будем перемещать конечную точку P'. Вместе с точкой P' будет перемещаться и луч PP', соединяющий ее с точкой P. В результате получится пучок лучей, исходящих в различных направлениях из точки P, как если бы она была источником света (рис. 72). Пусть $F'F'$ — волновой фронт, соответствующий этому пучку, проходящий через начальное положение точки P'. При смещении P' в точку Q', лежащую на луче $PA'Q'$, функция H получает приращение

$\Delta H = (PA'Q') - (PP')$, или, ввиду равенства оптических длин PP' и PA',

$$\Delta H = (PA'Q') - (PA') = (A'Q').$$

Если смещение $d\mathbf{r}' = \overrightarrow{P'Q'}$ бесконечно мало, то ΔH переходит в

$$dH = n'|d\mathbf{r}'|\cos\alpha' = n'(\mathbf{s}'\,d\mathbf{r}'),$$

где α' — угол между вектором $d\mathbf{r}'$ и единичным вектором луча $\mathbf{s}'$ в точке P', а n' — показатель преломления среды в этой точке. Аналогично найдется бесконечно малое приращение функции H, когда конечная точка P' остается неподвижной, а начальная точка P перемещается на $d\mathbf{r}$. В этом случае при вычислении надо воспользоваться пучком лучей, сходящихся в точке P' (рис. 73). В результате получится $dH = -n(\mathbf{s}\,d\mathbf{r})$, где n, $\mathbf{s}$ и $d\mathbf{r}$ имеют такой же смысл, что и в предыдущем случае, но относятся к начальной точке P. При изменении обоих аргументов $\mathbf{r}$ и $\mathbf{r}'$ получим

$$dH = n'(\mathbf{s}'\,d\mathbf{r}') - n(\mathbf{s}\,d\mathbf{r}). \tag{19.1}$$

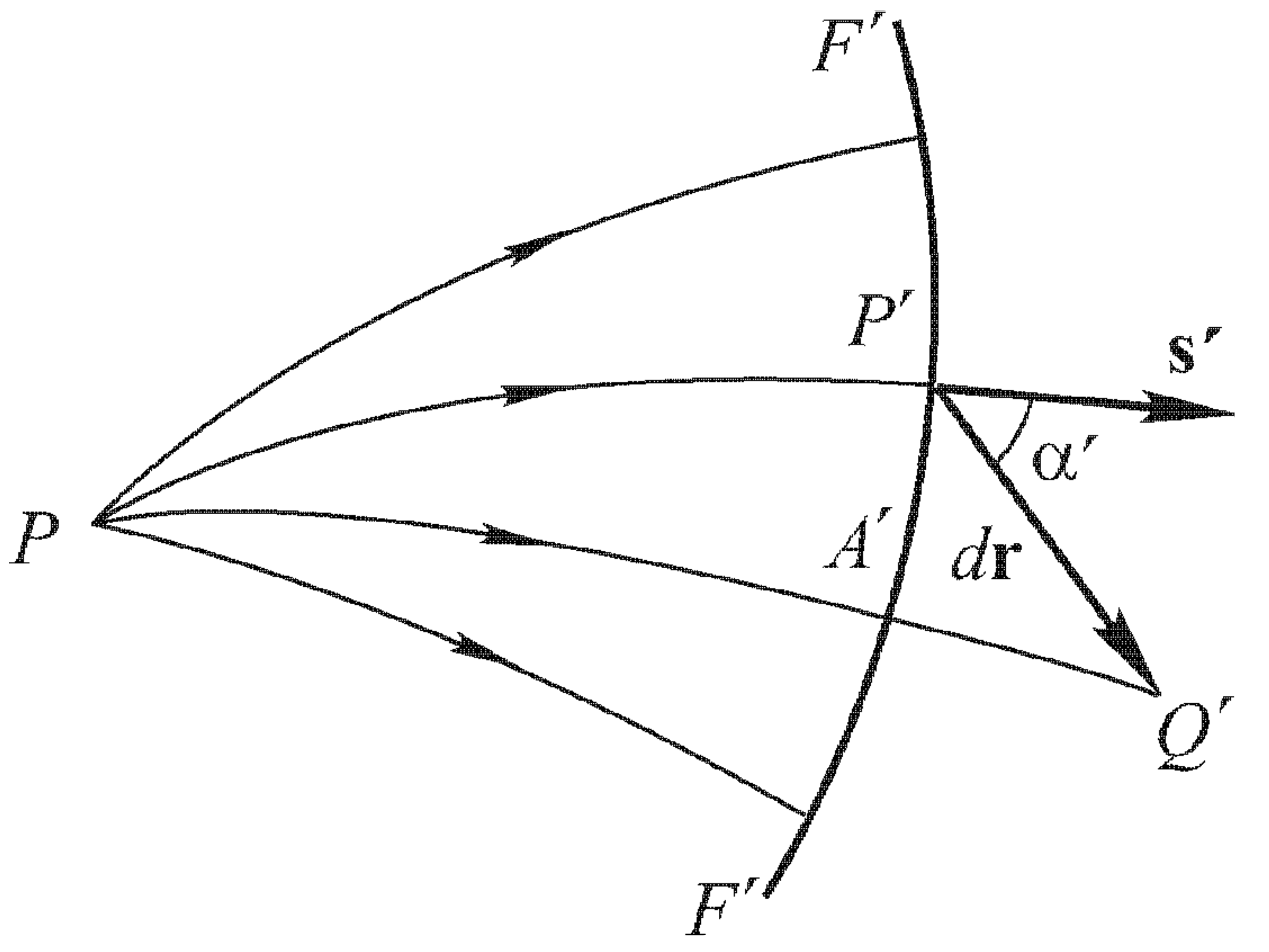


Рис. 72

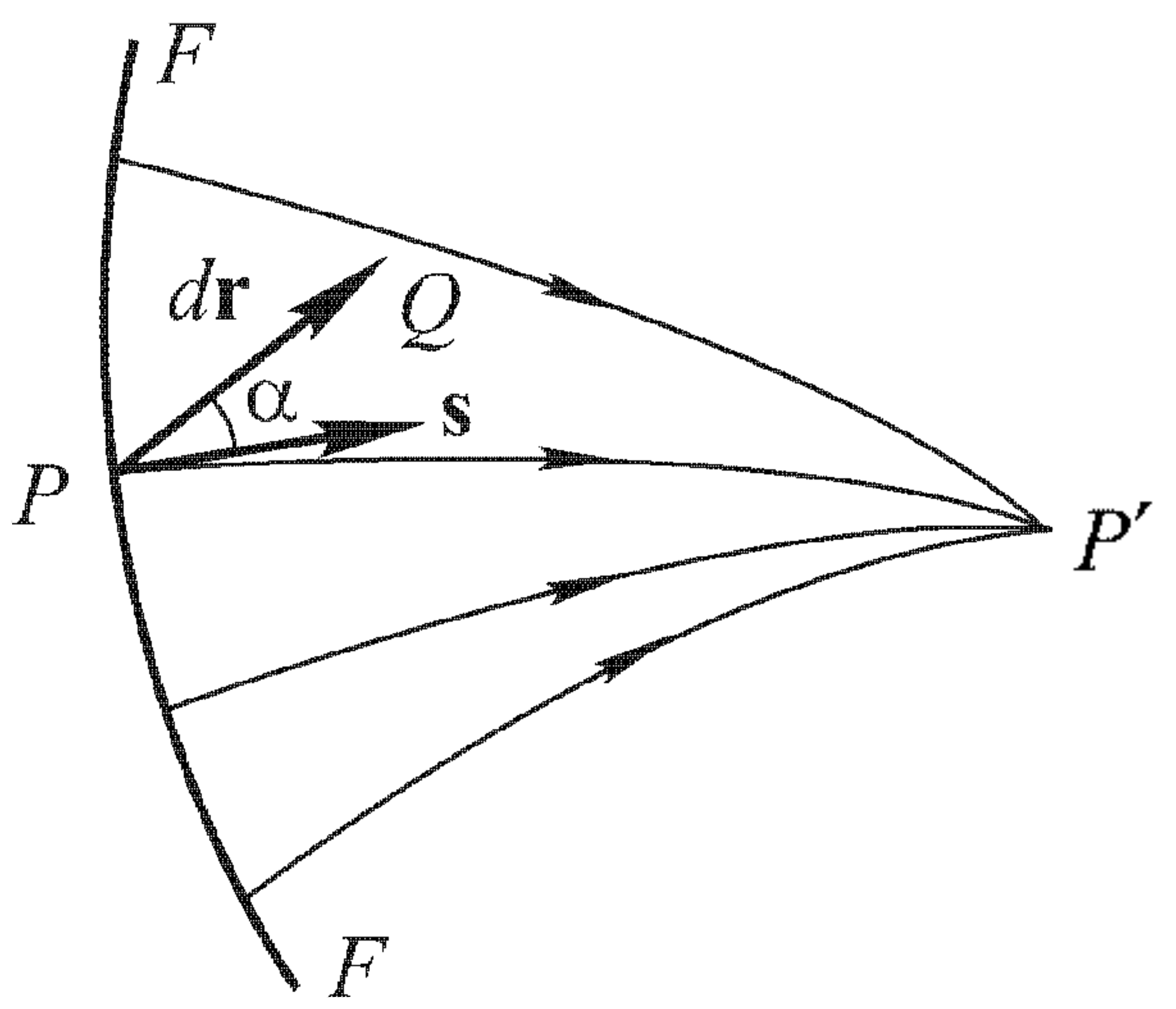


Рис. 73

Докажем теперь следующую теорему, называемую *теоремой косинусов*:

Пусть точка P' (рис. 74) *является стигматическим изображением точки* P. *Соединим эти точки произвольным лучом, направления которого в* P *и* P' *определяются единичными векторами* $\mathbf{s}$ *и* $\mathbf{s}'$. *Пусть* Q *и* Q' *— две точки, бесконечно близкие к* P *и* P'. *Для того чтобы точка* Q' *была стигматическим изображением точки* Q *в широких пучках лучей, необходимо и достаточно, чтобы разность*

$$n\mathbf{s}\,d\mathbf{l} - n'\mathbf{s}'\,d\mathbf{l}', \tag{19.2}$$

где $d\mathbf{l} = \overrightarrow{PQ}$, $d\mathbf{l}' = \overrightarrow{P'Q'}$, *не вависела от направления луча, соединяющего* P *с* P'.

Для доказательства необходимости теоремы допустим, что Q' является стигматическим изображением точки Q. Обозначим через H оптическую длину какого-либо луча, соединяющего сопряженные точки P и P', а через H' — оптическую длину луча, соединяющего сопряженные точки Q и Q'. В силу известного свойства сопряженных точек величины H и H', а с ними и разность $dH = H' - H = n'\mathbf{s}'\,d\mathbf{l}' - n\mathbf{s}\,d\mathbf{l}$ не зависят от направлений лучей, соединяющих P с P', а Q с Q'. Таким образом, выражение (19.2) не может зависеть от $\mathbf{s}$, и необходимость теоремы доказана.

Докажем теперь ее достаточность. Соединим точки Q и Q' лучом QQ'. Кроме того, через Q проведем произвольный луч QQ'' и отметим на нем такую точку Q'', чтобы оптическая длина (QQ'') была равна оптической длине (QQ'). Так как по условию теоремы P' является стигматическим изображением точки P, то из соображений непрерывности следует, что вектор $\overrightarrow{P'Q''} = d\mathbf{l}''$ бесконечно мал. В таком случае по формуле (19.1)

$$(QQ'') - (PP') = n'\mathbf{s}'\, d\mathbf{l}'' - n\mathbf{s}\, d\mathbf{l}.$$

По той же формуле

$$(QQ') - (PP') = n'\mathbf{s}'\, d\mathbf{l}' - n\mathbf{s}\, d\mathbf{l}.$$

Так как по построению $(QQ') = (QQ'')$, то $\mathbf{s}'\, d\mathbf{l}'' = \mathbf{s}'\, d\mathbf{l}'$, причем по условию теоремы это соотношение должно выполняться для всех направлений вектора $\mathbf{s}'$. Это возможно тогда и только тогда, когда $d\mathbf{l}'' = d\mathbf{l}'$, т.е. когда точка Q'' совпадает с Q'. Таким образом, любой луч, исходящий из Q, пройдет через Q', что и доказывает достаточность теоремы.

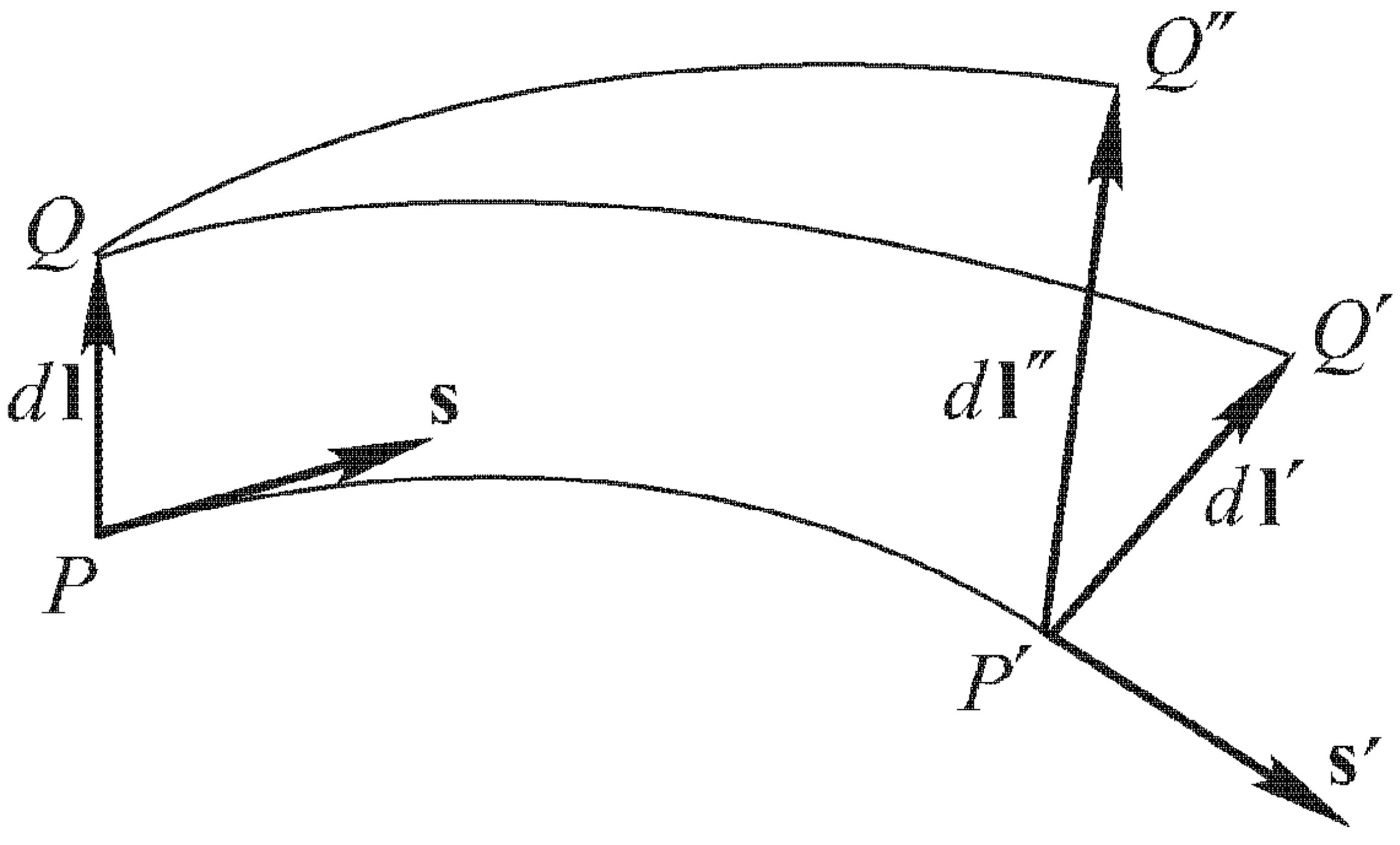


Рис. 74

Чтобы конечная кривая изображалась стигматически широкими пучками лучей, необходимо и достаточно, чтобы условия теоремы косинусов выполнялись для каждой пары сопряженных бесконечно малых отрезков этих кривых.

2. Условие синусов является следствием теоремы косинусов. Действительно, согласно этой теореме разность $n\mathbf{l}\mathbf{s} - n'\mathbf{l}'\mathbf{s}' = nl\sin u - n'l'\sin u'$ не должна зависеть от направления луча, соединяющего сопряженные точки P и P' (рис. 75). Но если луч идет вдоль оптической оси, то указанная разность обращается в нуль. В результате снова получается соотношение (18.1), и притом не только как необходимое, но и как *достаточное условие.*

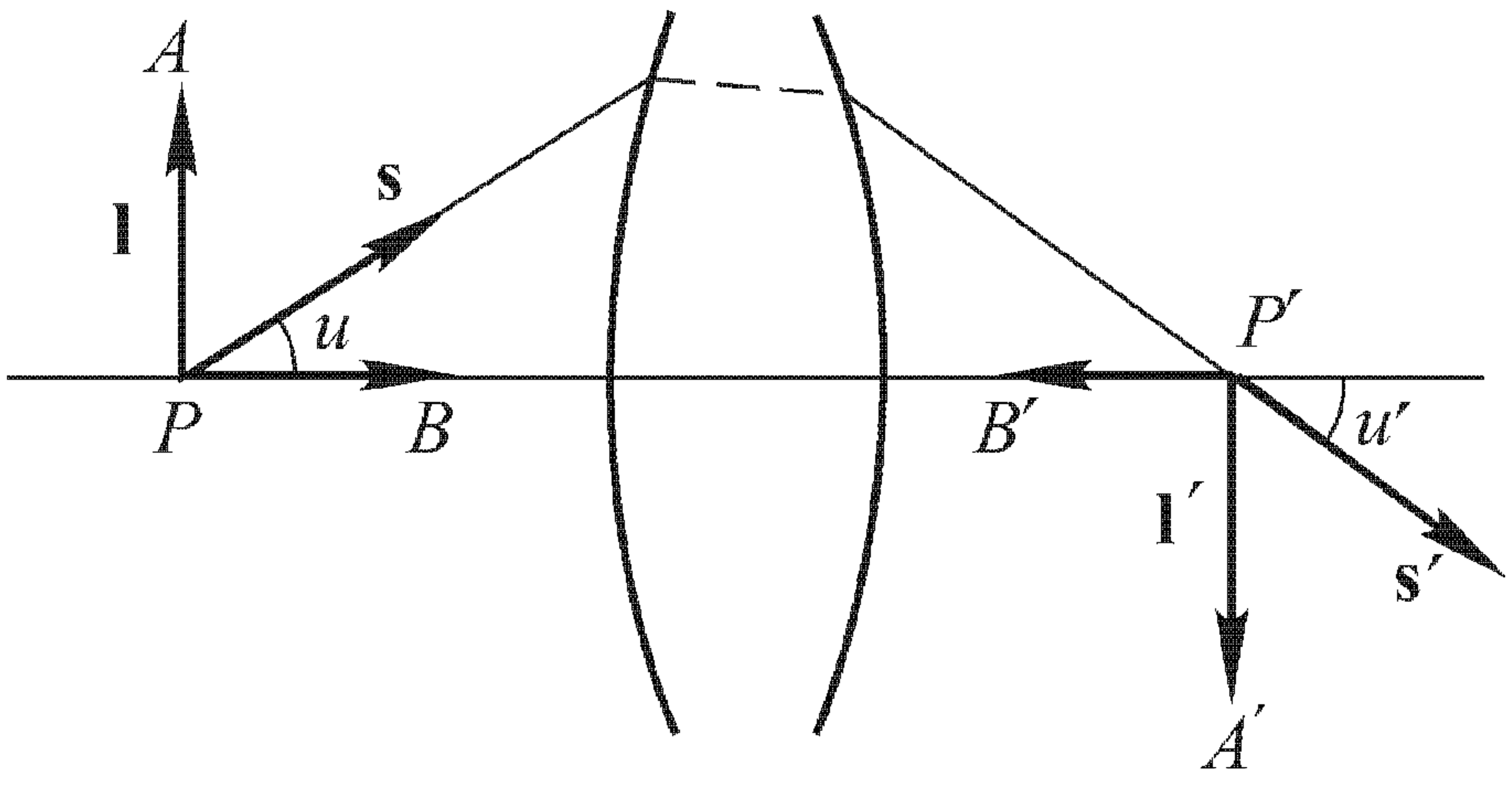


Рис. 75

Допустим теперь, что отрезок $l = PB$ лежит на оптической оси. Найдем условие, при котором он изображается стигматически широкими пучками лучей в виде отрезка $l' = P'B'$, также лежащего на оптической оси. В рассматриваемом случае при $u = 0$

$$n\mathbf{l}\mathbf{s} - n'\mathbf{l}'\mathbf{s}' = nl\cos u - n'l'\cos u' = nl - n'l'.$$

Поэтому на основании теоремы косинусов должно быть

$$nl\cos u - n'l'\cos u' = nl - n'l',$$

или

$$nl\sin^2(u/2) = n'l'\sin^2(u'/2), \tag{19.3}$$

каковы бы ни были значения углов u и u'. Это соотношение называется *условием Гершеля.*

Пусть P и P' — сопряженные апланатические точки. Если в бесконечно малых окрестностях этих точек существует другая пара апланатических точек Q и Q' (а следовательно, и бесконечное множество таких пар), то оптическая длина бесконечно малого линейного объекта, лежащего в окрестности точки P, будет равна оптической длине его изображения, получающегося в окрестности точки P'. Действительно, так как P и P' — апланатические точки, то по условию синусов отношение

$$\frac{\sin u}{\sin u'} = \frac{\cos(u/2)}{\cos(u'/2)} \cdot \frac{\sin(u/2)}{\sin(u'/2)}$$

не должно зависеть от u. С другой стороны, так как Q и Q' — также апланатические точки, то бесконечно малый отрезок PQ оптической оси изображается в виде отрезка $P'Q'$ той же оси. Поэтому должно соблюдаться и условие Гершеля: отношение

$$\frac{\sin^2(u/2)}{\sin^2(u'/2)}$$

не должно зависеть от угла u. Оба условия могут быть выполнены одновременно тогда и только тогда, когда $u = \pm u'$, т.е. когда апланатические точки P и P' являются узловыми или обратными узловыми точками системы. Но в таком случае, если отрезок l перпендикулярен к оптической оси, из (18.1) следует: $|nl| = |n'l'|$. Если же отрезок l лежит на оптической оси, то из (19.3) также следует $|nl| = |n'l'|$. Таким образом, теорема доказана для двух частных случаев: когда отрезок l лежит на оптической оси и когда он перпендикулярен к ней. Тем самым она доказана для отрезка l произвольного направления.

Во всех центрированных системах, применяющихся на практике, линейное увеличение, как правило, отлично от n/n'. Поэтому из доказанной теоремы следует, что для таких систем апланатические пары точек, если они существуют, могут быть лишь *изолированными точками* оптической оси. Это значит, что для каждой пары апланатических точек можно указать конечные интервалы оптической оси, содержащие эти точки, внутри которых нет другой пары апланатических точек.

В приближении параксиальной оптики условие синусов и условие Гершеля выполняются всегда. Первое из них в указанном приближении переходит в теорему Лагранжа–Гельмгольца. Для доказательства второго из соотношений (11.20), (11.17), (11.13), (10.6) находим

$$\frac{\delta X'}{\delta X} = \frac{l'}{l} = -\frac{f^2}{X^2}\frac{f'}{f} = \frac{n'}{n}\frac{y'^2}{y^2} = \frac{n}{n'}\frac{u^2}{u'^2},$$

а это при малых углах u и u' совпадает с условием Гершеля (19.3). Таким образом, равенство углов u и u', а следовательно, и равенство оптических длин nl и $n'l'$ не обязательны. Благодаря этому в параксиальной оптике поперечное и продольное увеличения могут принимать любые значения, не обязательно равные n/n'.

3. *Пусть P' является стигматическим изображением точки P. Для того чтобы бесконечно малый элемент плоскости S, проходящий через P, изображался стигматически в виде бесконечно малого элемента плоскости S', проходящей через P', необходимо и достаточно, чтобы выполнялось условие косинусов для двух бесконечно малых непараллельных отрезков, лежащих в плоскости S и проходящих через точку P.*

Необходимость теоремы очевидна. Для доказательства ее достаточности соединим P и P' произвольным лучом. Пусть $d\mathbf{l}_1$ и $d\mathbf{l}_2$ — два бесконечно малых неколлинеарных вектора, проходящих через точку P, для которых удовлетворяется условие теоремы косинусов. Таким образом, по условию разности

$$\begin{aligned} n'\mathbf{s}'\,d\mathbf{l}'_1 - n\mathbf{s}\,d\mathbf{l}_1 &= dH_1, \\ n'\mathbf{s}'\,d\mathbf{l}'_2 - n\mathbf{s}\,d\mathbf{l}_2 &= dH_2 \end{aligned} \tag{19.4}$$

не зависят от направления луча, соединяющего P с P'. Но они могут зависеть от направлений векторов $d\mathbf{l}_1$ и $d\mathbf{l}_2$. Произвольный вектор $d\mathbf{l}$, проходящий через точку P и лежащий в плоскости предмета, можно разложить по векторам $d\mathbf{l}_1$ и $d\mathbf{l}_2$:

$$d\mathbf{l} = a\,d\mathbf{l}_1 + b\,d\mathbf{l}_2,$$

причем коэффициенты a и b не зависят от $\mathbf{s}$. Введем вектор

$$d\mathbf{l}' = a\,d\mathbf{l}'_1 + b\,d\mathbf{l}'_2,$$

умножим соотношения (19.4) на a и b и сложим. Получим

$$n'\mathbf{s}'\,d\mathbf{l}' - n\mathbf{s}\,d\mathbf{l} = dH, \tag{19.5}$$

где $dH = a\,dH_1 + b\,dH_2$. Отсюда видно, что разность (19.5) не зависит от $\mathbf{s}$, т.е. условие теоремы косинусов выполняется для произвольного вектора $d\mathbf{l}$, проходящего через точку P и лежащего в плоскости предмета. Следовательно, предмет изобразится оптической системой стигматически. Вообще говоря, изображение не будет подобно самому предмету.

Направления произвольных неколлинеарных векторов $d\mathbf{l}_1$ и $d\mathbf{l}_2$ можно принять за координатные оси Y и Z в плоскости предмета, направления оптически сопряженных с ними отрезков $d\mathbf{l}'_1$ и $d\mathbf{l}'_2$ — за координатные оси Y' и Z' в плоскости изображения, а сами точки P и P' — за начала координат соответствующих координатных систем. Тогда в окрестностях точек P и P' координатные оси Y и Z изобразятся оптической системой координатными осями Y' и Z'. Координаты сопряженных точек, бесконечно близких к P и P', будут связаны линейными соотношениями

$$y' = Ay, \qquad z' = Bz. \tag{19.6}$$

Координатная система $Y'Z'$, вообще говоря, будет косоугольной, даже в том случае, когда система YZ прямоугольная. Однако если в плоскостях предмета и его изображения ввести прямоугольные системы координат, то из соотношений (19.6) и из формул преобразования координат непосредственно следует, что прямоугольные координаты сопряженных точек будут связаны формулами линейного преобразования

$$y' = A_{11}y + A_{12}z, \qquad z' = A_{21}y + A_{22}z. \tag{19.7}$$

Из свойств линейного преобразования следует, что в плоскости предмета существует пара взаимно перпендикулярных прямых, которой в плоскости изображения соответствует пара также взаимно перпендикулярных прямых. Если эти четыре прямые принять за координатные оси, то формулы преобразования снова примут вид (19.6), с тем отличием, что теперь обе координатные системы прямоугольны. Вообще говоря, $A \neq B$. Поэтому изображение бесконечно малой площадки происходит с нарушением подобия: бесконечно малый круг изображается в виде эллипса. Только в частном случае, когда $A = B$, система дает подобные изображения бесконечно малых площадок, находящихся в окрестности точки P.

4. Следуя Каратеодори (1873–1950), говорят, что *световой луч лежит в поле инструмента, если он действительно проходит через диафрагмы из пространства предметов в пространство изображений.* Говорят также, что *отрезок кривой лежит тангенциально в поле инструмента, если все лучи, касающиеся этого отрезка, лежат в поле инструмента.*

Пусть бесконечно малая площадка изображается оптической системой стигматически. Пусть, далее, $d\mathbf{l}_1$ *и* $d\mathbf{l}_2$ *— бесконечно малые не параллельные отрезки, пересекающиеся в пределах площадки и лежащие в ее плоскости. Если эти отрезки лежат тангенциально в поле инструмента, то рассматриваемая площадка изображается оптической системой с сохранением подобия. При этом оптическая длина любого отрезка, лежащего на площадке, равна оптической длине сопряженного с ним отрезка.*

По условию теоремы лучи, выходящие из P (рис. 76) в направлениях $d\mathbf{l}_1$ и $d\mathbf{l}_2$, лежат в поле инструмента. В пространстве изображений они пройдут

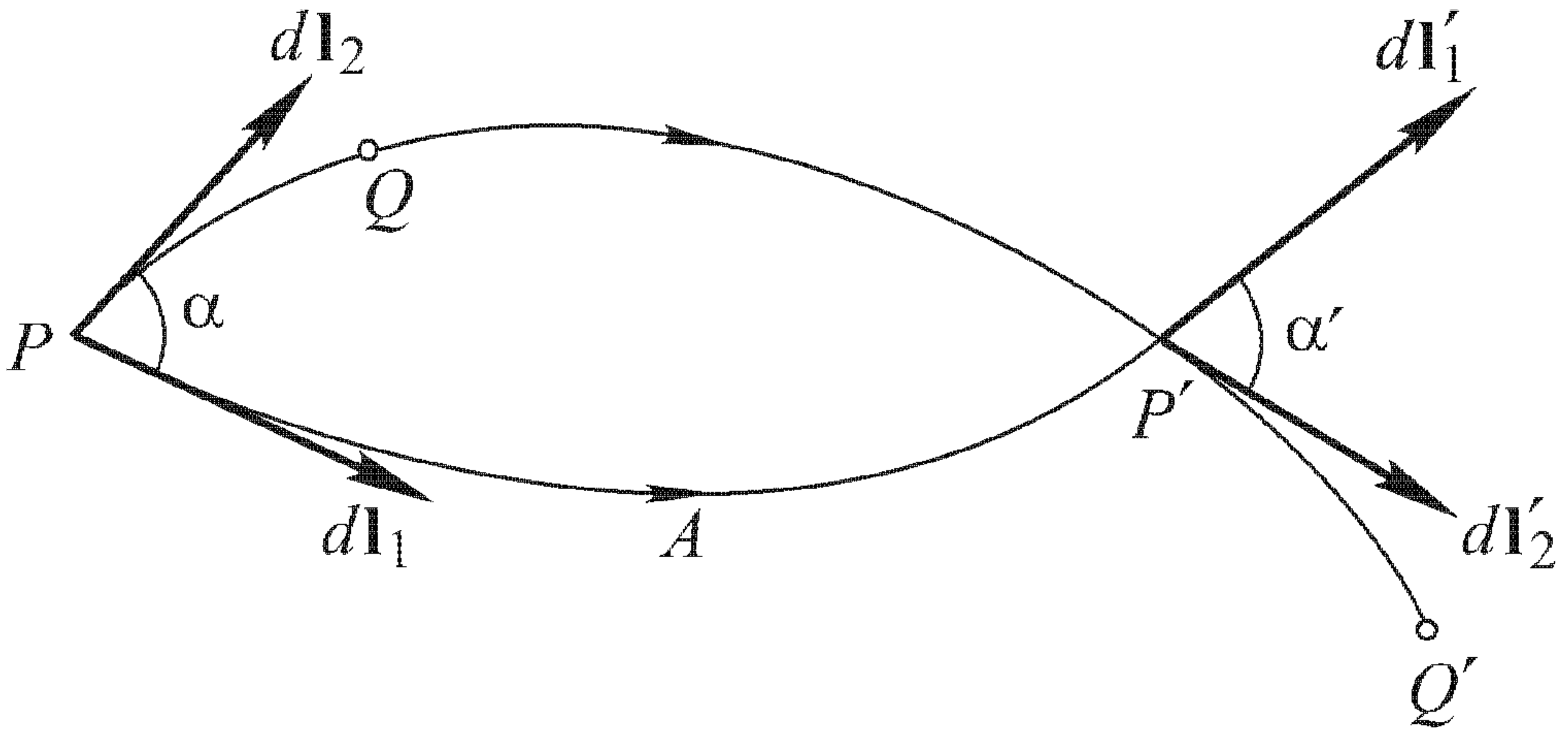


Рис. 76

в направлениях оптически сопряженных отрезков $d\mathbf{l}'_1$ и $d\mathbf{l}'_2$. Рассмотрим сначала изображение отрезка $d\mathbf{l}_1$. Возьмем два луча, исходящих из P в направлениях $d\mathbf{l}_2$ и $d\mathbf{l}_1$. На основании теоремы косинусов

$$n\,dl_1 \cos\alpha - n'\,dl'_1 \cos\alpha' = n\,dl_1 - n'\,dl'_1. \tag{19.8}$$

Аналогично, рассматривая изображение отрезка $d\mathbf{l}_2$,

$$n\,dl_2 \cos\alpha - n'\,dl'_2 \cos\alpha' = n\,dl_2 - n'\,dl'_2. \tag{19.9}$$

Допустим, что $dl_1 = dl_2$. Докажем, что тогда $dl'_1 = dl'_2$. В самом деле, почленное вычитание (19.9) из (19.8) дает

$$n'\,dl'_1(1-\cos\alpha') = n'\,dl'_2(1-\cos\alpha'). \tag{19.10}$$

Разность $1-\cos\alpha'$ не может обращаться в нуль. Действительно, по предположению угол α не равен нулю. Но заданием направления в какой-либо точке луч определяется однозначно. Если бы $\alpha' = 0$, т.е. направления обоих рассматриваемых нами лучей в точке P' совпадали, то они совпадали бы и во всех других точках, в частности в начальной точке P. Значит, было бы $\alpha = 0$, вопреки предположению. Поэтому, сокращая на $n'(1-\cos\alpha')$, из (19.10) находим $dl'_1 = dl'_2$, что и требовалось доказать.

Из доказанного следует, что отрезки $d\mathbf{l}_1$ и $d\mathbf{l}_2$ изображаются оптической системой с одинаковым увеличением. Следовательно, в рассматриваемом случае в формулах (19.6) $A = B$, т.е. увеличение любого отрезка в плоскости предмета не зависит от его направления. Отсюда следует, что изображение происходит с сохранением подобия, т.е. является *конформным*.

Но изображение с сохранением подобия характеризуется также сохранением углов. Следовательно, $\alpha = \alpha'$, и формула (19.8) дает $n\,dl_1 = n'\,dl_1'$. Вообще, для всякого отрезка dl, лежащего в плоскости предмета, $n\,dl = n'\,dl'$, $dl'/dl = n/n'$, и вторая часть теоремы доказана.

Итак, *стигматические изображения площадок, тангенциально лежащих в поле инструмента, могут происходить только с вполне определенным увеличением* n/n'. В частности, когда показатели преломления пространств предметов и изображений одинаковы, увеличение равно единице. Это утверждение перестает быть справедливым для площадок, не лежащих тангенциально в поле инструмента.

Примером может служить преломление на сферической поверхности (см. рис. 67). Сфера S отображается на сферу S' стигматически широкими пучками лучей. Однако линейное увеличение, как видно из построения, равно отношению квадратов показателей преломления, а не их первых степеней. Причина этого в том, что ни одна из сфер S и S' не лежит тангенциально в поле инструмента. Напротив, если линейный объект поместить в точку O, то, поскольку последняя является парой совпадающих узловых точек, линейное увеличение будет равно просто отношению показателей преломления в согласии с обсуждаемой нами общей теоремой. Действительно, ввиду шаровой симметрии любой линейный объект, помещенный в центре O, лежит тангенциально в поле инструмента.

Бесконечно малую часть конечной поверхности можно рассматривать как бесконечно малую плоскую площадку. Поэтому *для стигматического изображения конечной поверхности необходимо и достаточно, чтобы стигматически изображались все бесконечно малые площадки, на которые можно разбить эту поверхность.*

5. Рассмотрим, наконец, стигматические изображения объемных объектов широкими пучками лучей. Этот вопрос может быть исследован в точности так же, как и аналогичный вопрос для поверхностных объектов. В частности, может быть доказана следующая теорема:

Пусть точка P' является стигматическим изображением точки P. Для того чтобы бесконечно малый элемент объема в окрестности точки P изображался стигматически, необходимо и достаточно, чтобы выполнялось условие теоремы косинусов для трех бесконечно малых отрезков, проходящих через точку P и не лежащих в одной плоскости.

Новым по сравнению с изображениями элементов поверхностей является то, что в случае стигматического изображения элементов объема всегда существуют три отрезка dl_1, dl_2, dl_3, не находящиеся в одной плоскости и лежащие тангенциально в поле инструмента. Поэтому, повторяя рассуждения, приведенные применительно к изображениям элементов поверхности, приходим к заключению, что эти три отрезка изображаются с одним и тем же увеличением. Как следствие этого, получаем следующую теорему:

Стигматическое изображение элементов объема, если оно осуществляется широкими пучками лучей, всегда конформно, т.е. происходит с сохранением подобия. При этом линейное увеличение равно n/n', так что оптическая длина предмета всегда равна оптической длине изображения.

§ 20. Об абсолютных оптических инструментах

1. С точки зрения геометрической оптики идеалом был бы оптический инструмент, изображающий стигматически широкими пучками лучей каждую точку пространства предметов в виде точки пространства изображений. Такой инструмент называется *абсолютным.* Обсудим вопрос о принципиальной возможности абсолютных оптических инструментов и исследуем их свойства.

Как доказано в предыдущем параграфе, изображения бесконечно малых объектов, даваемые абсолютным оптическим инструментом, всегда *конформны.* При этом оптическая длина любой линии равна оптической длине ее изображения. Отсюда следует, что *в абсолютном оптическом инструменте оптическая длина луча между двумя сопряженными точками одинакова для всех пар сопряженных точек.* Это положение называется *теоремой Каратеодори.*

Для доказательства возьмем две пары сопряженных точек: P, P' и Q, Q' (рис. 76). Через точку P проведем пучок лучей во всевозможных направлениях. Все эти лучи пересекутся в точке P'. Один из них пройдет через точку Q. Возьмем другой произвольный луч PAP'. В силу свойства сопряженных точек $(PAP') = (PQP')$. Луч PQP', поскольку он проходит через точку Q, должен пройти и через сопряженную ей точку Q'. При этом кривая $P'Q'$ будет оптическим изображением кривой PQ, а потому $(PQ) = (P'Q')$. Отсюда следует $(QP'Q') = (PQP')$. Сравнивая это равенство с предыдущим, получаем: $(PAP') = (QP'Q')$, что и требовалось доказать.

2. Допустим теперь, что пространства предметов и изображений *однородны,* т.е. показатели преломления n и n' не зависят от координат. В этом случае *любая прямая пространства предметов изобразится абсолютным инструментом в виде прямой пространства изображений.* В самом деле, проведем через произвольную точку O пространства предметов три луча, лежащих в поле инструмента и не находящихся в одной плоскости. В пространстве изображений сопряженные с ними лучи пересекутся в сопряженной точке O'. Примем эти лучи за координатные оси в пространствах предметов и изображений. Ясно, что координатные оси одной координатной системы будут изображаться координатными осями другой координатной системы. При этом без потери общности можно принять, что одноименные оси являются сопряженными. Так как изображение происходит с постоянным увеличением n/n', то при надлежащем выборе положительных направлений координатных осей координаты сопряженных точек будут связаны соотношениями

$$x' = \frac{n}{n'}x, \quad y' = \frac{n}{n'}y, \quad z' = \frac{n}{n'}z, \tag{20.1}$$

справедливыми для всех точек пространства предметов и пространства изображений, а не только для окрестностей сопряженных точек O и O'. Из них непосредственно следует, что любая прямая изображается прямой. Таким образом, *при постоянных n и n' абсолютный инструмент является телескопической системой.*

Докажем теперь, что *при постоянных n и n' абсолютный оптический инструмент возможен только при $n = n'$.* Для доказательства заметим, что линейное увеличение n/n' не зависит от положения объекта. Чтобы определить это увеличение, поместим бесконечно малый объект тангенциально на границе раздела сред. Тогда его изображение, возникающее как при преломлении, так и при отражении, совместится с самим объектом. Отсюда следует, что $n/n' = 1$, и наше утверждение доказано. Таким образом, показатели преломления сред должны быть одинаковы, и никакого преломления не будет.

Остается исследовать только отражение. По теореме Каратеодори оптическая длина луча между сопряженными точками не зависит от положения объекта. Поместим точечный объект на границе раздела сред. Его изображение получится в той же точке. Поэтому оптическая длина луча между объектом и его изображением должна равняться нулю, независимо от положения объекта. При преломлении, когда оптические длины лучей на всех участках сохраняют знаки, это было бы возможно только тогда, когда изображение совпадает с самим объектом. Но в этом случае, как мы видели, никакого преломления нет, и об изображении можно говорить лишь условно. Однако при отражении, если изображение получается мнимым, оптические длины лучей на разных участках могут иметь противоположные знаки. Тогда возможен и нетривиальный случай, когда объект и его изображение находятся в разных местах, хотя оптические длины лучей между ними и обращаются в нуль. При этом формулы (20.1) переходят в

$$x' = x, \quad y' = y, \quad z' = z. \tag{20.2}$$

Такой случай осуществляется в плоском зеркале или в системе плоских зеркал. Это единственный абсолютный оптический инструмент, возможный при постоянных n и n'.

3. *Нетривиальные абсолютные оптические инструменты, дающие изображения с увеличением, отличным от единицы, возможны только с использованием неоднородных сред и криволинейных лучей.* Первый пример такого инструмента был приведен Максвеллом. В дальнейшем были приведены и другие примеры. Максвелл назвал свой инструмент *«рыбьим глазом»*, хотя никакого сходства с глазом рыбы у него нет. Более того, «рыбий глаз» Максвелла вообще трудно назвать инструментом в обычном смысле этого слова. Действительно, он представляет собой неограниченную неоднородную среду, показатель преломления которой меняется в пространстве таким образом, что любой луч в этой среде имеет форму окружности. К идее «рыбьего глаза» естественнее всего прийти с помощью геометрического построения, называемого *стереографической проекцией.*

Такая проекция осуществляется следующим образом. Возьмем сферу S, через центр которой проведем координатную плоскость XY (рис. 77). Пусть

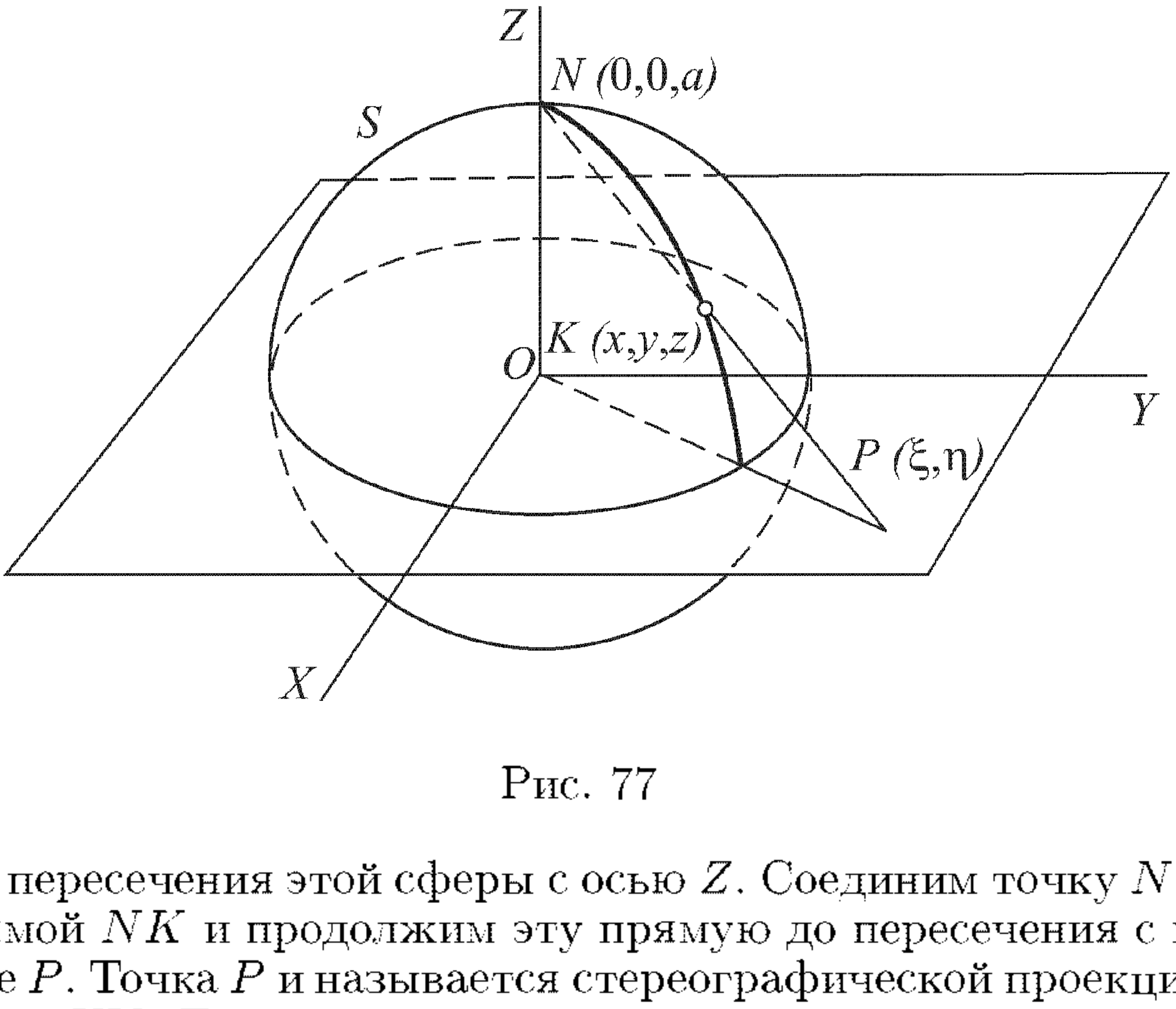


Рис. 77

N — точка пересечения этой сферы с осью Z. Соединим точку N с точкой K сферы прямой NK и продолжим эту прямую до пересечения с плоскостью XY в точке P. Точка P и называется стереографической проекцией точки K на плоскость XY. Легко показать, что координаты x, y, z точки K связаны с координатами ξ, η соотношениями

$$\xi = a\frac{x}{a-z}, \qquad \eta = a\frac{y}{a-z}, \tag{20.3}$$

$$x = 2a^2\frac{\xi}{\xi^2+\eta^2+a^2}, \qquad y = 2a^2\frac{\eta}{\xi^2+\eta^2+a^2}, \qquad z = a\,\frac{\xi^2+\eta^2-a^2}{\xi^2+\eta^2+a^2}, \tag{20.4}$$

где a — радиус сферы S.

Каждая окружность сферы S при стереографической проекции преобразуется в окружность плоскости XY и обратно. В самом деле, плоскость $Ax+By+Cz+D=0$ пересекает сферу S по окружности. Если в уравнение этой плоскости подставить значения координат (20.4), то получится уравнение второй степени, в котором будет отсутствовать член с произведением $\xi\eta$, а квадраты ξ^2 и η^2 войдут только в комбинации $\xi^2+\eta^2$. Такое уравнение есть уравнение окружности.

Пусть $d\sigma = \sqrt{dx^2+dy^2+dz^2}$ — элемент длины на сфере S. При стереографической проекции он отобразится элементом длины $ds = \sqrt{d\xi^2+d\eta^2}$ на плоскости XY. Пользуясь формулами (20.4), нетрудно показать, что эти элементы связаны соотношением

$$d\sigma = \frac{2}{1+(r/a)^2}\,ds, \tag{20.5}$$

где $r^2 = \xi^2+\eta^2$. Кривая кратчайшей длины, проведенная на какой-либо поверхности между двумя точками, называется *геодезической линией* этой поверхности. Для сферы геодезическими линиями являются дуги больших кругов. Возьмем на сфере S любые диаметрально противоположные точки $K(x_0, y_0, z_0)$ и $K'(-x_0, -y_0, -z_0)$. Пусть P и P' — их стереографические проекции на плоскость XY. Через прямую, соединяющую точки K и K', можно провести бесконечное множество плоскостей, пересекающих сферу S вдоль больших кругов, проходящих через эти точки. При стереографической проекции эти большие круги изобразятся окружностями, соединяющими точки P и P' (рис. 78). Вдоль этих окружностей интеграл $\int ds/(1+r^2/a^2)$ принимает одно и то же значение, как это непосредственно видно из соотношения (20.5).

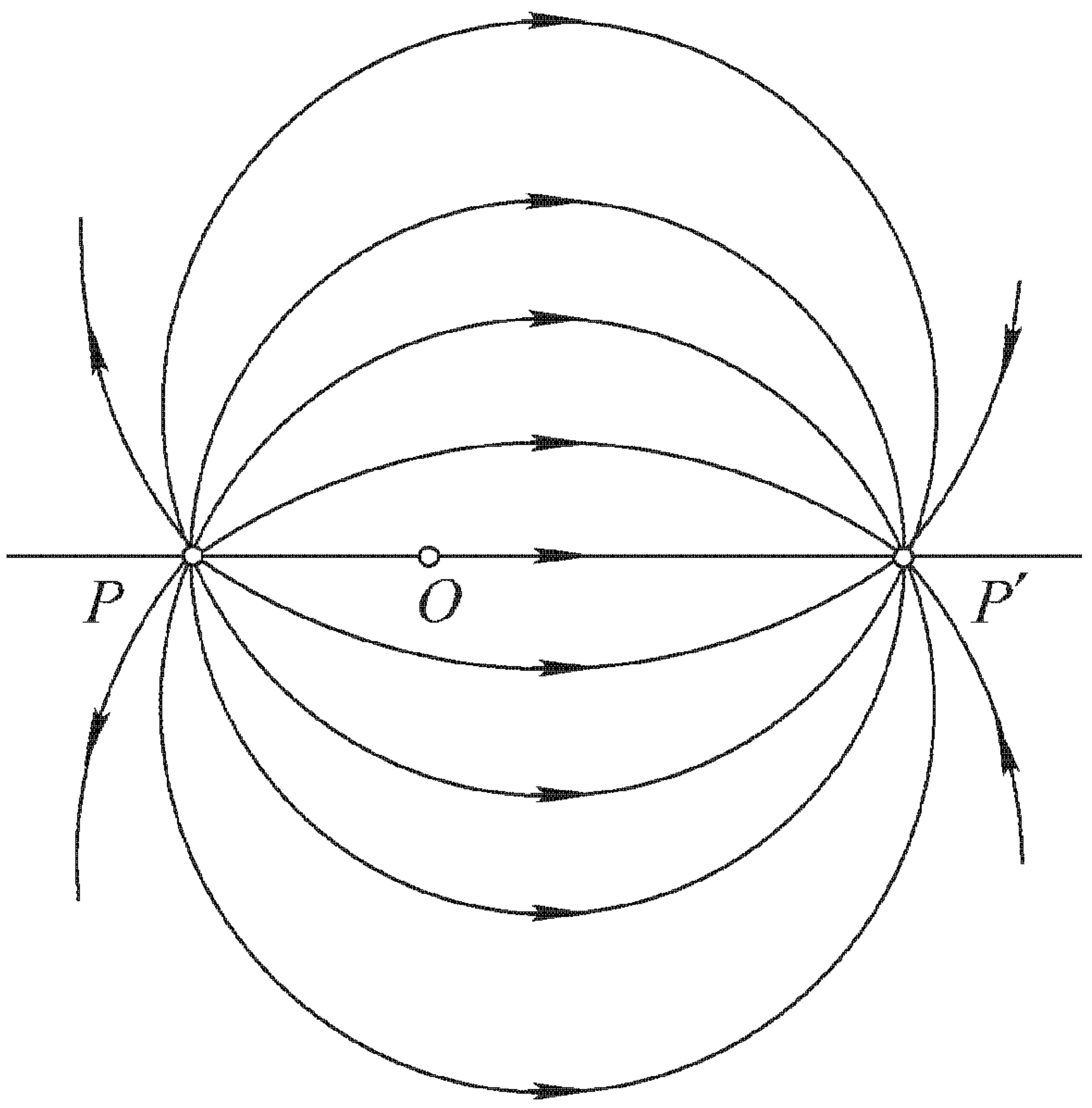


Рис. 78

Вообразим теперь сферически симметричную среду с центром в начале координат O, показатель преломления которой определяется выражением

$$n = \frac{n_0}{1+r^2/a^2}, \tag{20.6}$$

где n_0 — постоянная, а r — расстояние от начала координат. Тогда оптические длины всех окружностей между точками P и P', изображенных на рис. 78, будут одинаковыми. Отсюда следует, что все эти окружности будут совпадать со световыми лучами, исходящими из точки P и сходящимися в точке P'. Одна из этих окружностей вырождается в прямую, соединяющую точки P и P'. Она получается в результате стереографической проекции большого круга, лежащего в плоскости NKK'. Начало координат O лежит в той же плоскости, а следовательно, и на прямой PP'. Вообще говоря, его положение не совпадает с серединой отрезка PP'. Исключение составляет только случай, когда диаметрально противоположные точки K и K' сферы S лежат в координатной плоскости XY. В этом случае большой круг, лежащий в той же плоскости, совпадает со своей стереографической проекцией, а потому радиус кривизны светового луча будет равен радиусу сферы S, т.е. a.

Ввиду шаровой симметрии все плоскости, проходящие через центр O, эквивалентны. Поэтому луч, исходящий из точки P, пройдет через сопряженную точку P', независимо от того, лежит он в координатной плоскости XY или не лежит. Кроме того, приведенные нами рассуждения применимы для любой пары диаметрально противоположных точек K и K' и соответствующей им пары сопряженных точек P и P'.

Таким образом, получается следующий результат. Все лучи, выходящие из любой точки среды P, описав окружности, собираются в другой точке P', т.е. точка P' является стигматическим изображением точки P в широких пучках лучей. Иными словами, среда с показателем преломления (20.6) обладает свойствами абсолютного оптического инструмента. Она и была названа Максвеллом «рыбьим глазом».

§ 21. Глаз и зрение

1. На рис. 79 приведен горизонтальный разрез правого глаза человека (вид сверху). Глазное яблоко заключено в плотную белую непрозрачную оболочку ww, называемую *склерой.* Через заднюю стенку склеры проходит глазной нерв e и кровеносные сосуды. Передняя часть склеры переходит в прозрачную *роговую оболочку*, или *роговицу* h, более выпуклую, чем склера. Толщина склеры от 0,4 до 1,1 мм, роговицы — около 0,5 мм. За роговицей следуют *передняя глазная камера* a, *хрусталик* l, *задняя глазная камера* gl. Эти части вместе с роговицей составляют *оптическую систему глаза*, дающую при преломлении лучей оптические изображения предметов. Толщина передней глазной камеры, а также хрусталика около 3,6 мм. Передняя глазная камера заполнена прозрачной жидкостью, называемой *водянистой влагой*, задняя — прозрачным студенистым веществом, называемым *стекловидным телом.* Показатели преломления обеих этих сред практически одинаковы и равны 1,336. Показатель преломления роговой оболочки равен 1,376.

Внутренняя поверхность склеры покрыта *сосудистой оболочкой* (толщиной около 0,4 мм), содержащей разветвления кровеносных сосудов, питающих глаз. Спереди она переходит в утолщенную часть st, содержащую кольцевидную ресничную мышцу uu, и далее в *радужную оболочку* ii, не одинаково окрашенную у различных людей и определяющую «цвет глаз». Посередине радужной оболочки имеется круглое отверстие, называемое *зрачком глаза*, которое выполняет роль апертурной диафрагмы. Изображение этого отверстия, даваемое расположенными перед ним водянистой влагой и роговицей, играет роль входного, а изображение, даваемое хрусталиком и стекловидным телом, — выходного зрачков глаза. Оба они практически совпадают с настоящим зрачком. В зависимости от яркости света зрачок может рефлекторно сужаться и расширяться посредством мышцы *uu*, с которой связана радужная оболочка. Диаметр его может меняться примерно от 2 до 8 мм. Этим достигается ограничение поперечных размеров пучков света, поступающих в глаз.

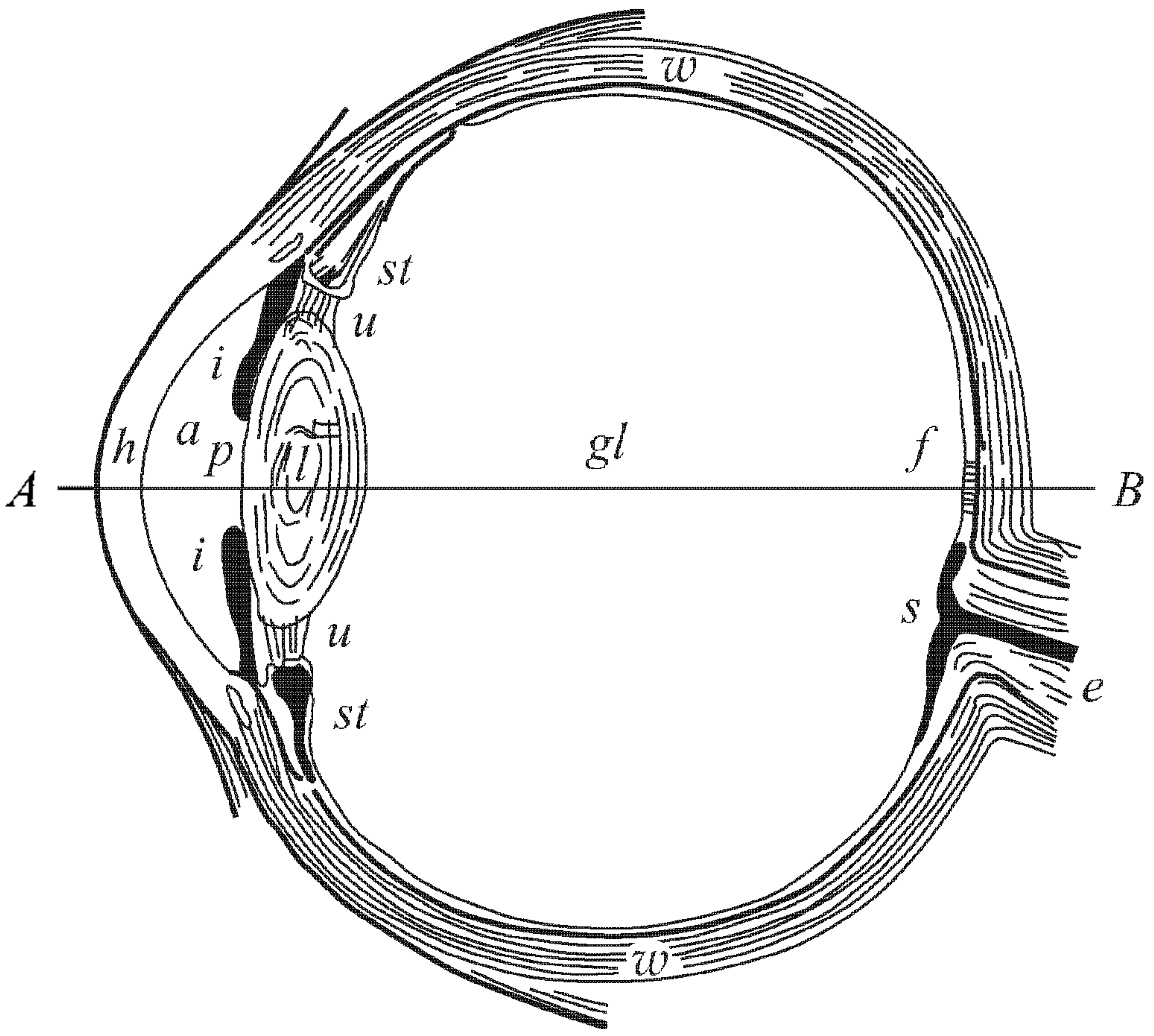


Рис. 79

2. На внутренней поверхности сосудистой оболочки расположена светочувствительная *сетчатая оболочка*, или *сетчатка* (*ретина*), состоящая главным образом из радиально расходящихся разветвлений зрительного нерва e, входящего во внутреннюю полость глаза несколько сбоку от линии прямого зрения AB, ближе к носовой полости. Около точки s толщина сетчатки составляет 0,2–0,22 мм; к периферии она непрерывно уменьшается. Оптическая система глаза дает изображение предметов на сетчатке, где оно вызывает зрительные раздражения, передаваемые в мозг по нервным волокнам.

Недалеко от точки s, где начинаются разветвления зрительного нерва, со стороны виска, находится так называемое *желтое пятно* f (горизонтальный поперечник 1–3 мм, вертикальный 0,8 мм). В его центральной части (размером $\approx 0{,}3$ мм по горизонтали и $\approx 0{,}2$ мм по вертикали) имеется углубление, называемое *центральной ямкой.* Здесь толщина сетчатки не превышает 0,08–0,1 мм. Желтое пятно и в особенности центральная ямка являются наиболее чувствительными местами сетчатки (при дневном зрении). Изображение рассматриваемой точки пространства всегда приводится к середине центральной ямки. Прямая, проходящая через эту точку и заднюю узловую точку глаза, вместе с параллельной ей прямой, проходящей через переднюю узловую точку, определяет так называемую *линию прямого зрения.* Она не совсем совпадает с *осью симметрии глаза*, так как преломляющие поверхности глаза не вполне симметричны и центрированы. Ось симметрии глаза можно провести в некотором смысле только условно. Обычно ее проводят через крайнюю, наиболее выступающую точку выпуклой поверхности роговицы (несколько выше точки A на рис. 79) и через центр зрачка. Она встречает заднюю стенку глаза где-то между точками f и s, образуя с линией прямого зрения AB угол около 5°.

3. Хрусталик l представляет собой прозрачное бесцветное тело, напоминающее двояковыпуклую линзу, передняя поверхность которой менее выпукла, чем задняя. Он состоит из слоев различной плотности, имеющих волокнистое строение. Наружный слой мягкий и почти студенистый; ядро же более твердое и вполне упругое. Показатель преломления наружного слоя хрусталика 1,405, средних слоев — около 1,429, ядра — 1,454. Мышца *ии*, рефлекторно напрягаясь или расслабляясь, может менять кривизну его поверхностей, главным образом передней. Этим осуществляется *аккомодация*, т.е. изменение оптической силы глаза, позволяющее фокусировать изображения на сетчатке. Так как деформация хрусталика может происходить только в определенных пределах, то для всякого глаза существуют определенные границы, в пределах которых глаз может отчетливо видеть предметы. Эти границы определяют так называемую *область аккомодации глаза.* Наиболее отдаленная граница, которую глаз может видеть отчетливо при вполне ослабленной мышце, называется *дальней*, а ближайшая граница, которую он способен отчетливо видеть при максимальном напряжении мышцы, — *ближней точкой ясного видения.* В ненапряженном состоянии нормальный глаз аккомодирован на рассмотрение бесконечно удаленных предметов, т.е. он собирает параллельные лучи в точке сетчатки. Таким образом, дальняя точка ясного видения для нормального глаза находится в бесконечности. В возрасте до десяти лет ближняя точка нормального глаза лежит впереди на расстоянии 7–8 см от глаза. К 30 годам это расстояние увеличивается примерно до 15 см, к 40–45 годам — до 25 см.

Приближая рассматриваемый предмет к глазу, мы увеличиваем *угол зрения*, а с ним и размеры изображения на сетчатке. Это позволяет рассмотреть более мелкие детали. Однако при максимально возможном приближении усиливается напряжение мышцы, деформирующей хрусталик. Работа глаза становится утомительной. В случае *нормального глаза* оптимальное расстояние для чтения и письма составляет около 25 см. Это расстояние для нормального глаза и принимается условно за *расстояние ясного зрения*. При необходимости рассмотреть (в течение ограниченного времени) более мелкие детали его можно уменьшить, оставаясь в пределах области аккомодации глаза.

4. У *близорукого глаза* оба предела ясного видения находятся ближе, а ближний предел *дальнозоркого глаза* — дальше, чем у нормального глаза. Дальний предел дальнозоркого глаза всегда отрицателен, т.е. дальняя точка ясного видения находится не спереди, а позади глаза. Такой глаз в ненапряженном состоянии может собирать на сетчатке только сходящиеся пучки лучей. При сильной дальнозоркости и ближний предел ясного зрения может сделаться отрицательным. У близорукого глаза в ненапряженном состоянии параллельные лучи сходятся перед, а у дальнозоркого — за сетчаткой.

Близорукость может быть обусловлена большей, а дальнозоркость — меньшей длиной глаза по сравнению с длиной нормального глаза. Отступления длины глаза от нормальной даже в пределах одного миллиметра уже могут приводить к заметной близорукости или дальнозоркости. Эти, а также другие недостатки глаза могут возникать из-за неправильных значений кривизны преломляющих поверхностей хрусталика и роговицы, несимметричности этих поверхностей, неправильного положения хрусталика и пр.

Термин «дальнозоркий» очень неудачен, так как с ним ассоциируется ошибочное представление, что дальнозоркость якобы способствует отчетливо видеть далекие предметы. Никакими преимуществами перед нормальным глазом дальнозоркий глаз не обладает, так как вся область перед глазом, отчетливо видимая дальнозорким глазом, отчетливо видна и нормальным глазом. Но область между ближними точками ясного видения дальнозоркого и нормального глаза недоступна для отчетливого рассматривания дальнозорким глазом. Близорукому глазу, напротив, недоступна область, лежащая за дальней точкой ясного видения нормального глаза ¹).

---

¹) У близорукого глаза есть по крайней мере то преимущество, что он может выявить более мелкие детали, приближая рассматриваемый предмет на меньшие расстояния, чем это возможно для нормального глаза. У дальнозоркого глаза аналогичное преимущество появилось бы, если бы возникла необходимость рассматривать мнимые изображения, образующиеся с задней стороны глаза.

Как уже отмечено выше, с возрастом (в основном из-за уплотнения хрусталика, теряющего способность достаточно сжиматься) ближняя точка ясного видения удаляется от глаза. Для нормального глаза это удаление к 50 годам составляет около 50 см. Читать на таком расстоянии уже трудно. В дальнейшем указанная точка уходит еще дальше, удаляется в бесконечность и к 60–65 годам перескакивает на противоположную сторону глаза. Затем она начинает приближаться к нему, но уже с противоположной стороны. Это явление называется *старческой дальнозоркостью.*

Дальнозоркость устраняется применением очков с *положительными* (т.е. собирающими), а близорукость — с *отрицательными* (т.е. рассеивающими) стеклами. Положительные стекла приближают, а отрицательные отдаляют ближнюю точку ясного видения.

5. Систематические измерения привели к установлению средних значений всех параметров нормального глаза. Совокупность этих значений определяет так называемый *схематический глаз.*

В табл. 3 приведены эти параметры, как они были определены Гульстрандом. Так как показатели преломления пространств предметов ($n = 1$) и изображений ($n' = 1{,}336$) различны, то переднее f и заднее f' фокусные расстояния не совпадают по величине ($f'/f = -1{,}336$). Задняя узловая точка K' расположена внутри глаза на расстоянии f от заднего фокуса F', т.е. от сетчатой оболочки (см. рис. 43). Если линейный размер предмета равен l, то он виден из точки нахождения глаза под углом зрения $\alpha = l/L$. Поэтому линейный размер изображения на сетчатке будет $l' = f\alpha = fl/L$, или при рассматривании удаленных предметов $\approx 17\alpha$ мм.

6. Сетчатка глаза имеет весьма сложное строение. Схематический разрез ее приведен на рис. 80 *а* (предполагается, что свет падает сверху). Сетчатка состоит из нескольких слоев нервных клеток различного типа и назначения. В последнем слое находятся *светочувствительные рецепторные клетки — палочки и колбочки* (рис. 80 *б*). В них свет вызывает *первичные раздражения*, превращающиеся в электрические импульсы, которые по волокнам зрительного нерва передаются в кору головного мозга. Общая длина палочки 0,06–0,08 мм, колбочки — около 0,035 мм. Число колбочек в глазу достигает 7 миллионов, палочек — до 130 миллионов. Периферия сетчатки заполнена почти исключительно палочками. Число колбочек на единицу площади возрастает по мере приближения к желтому пятну, достигая максимума в его середине. Число колбочек на площади желтого пятна 13–15 тысяч. В его центральной части, в пределах кру-

Таблица 3

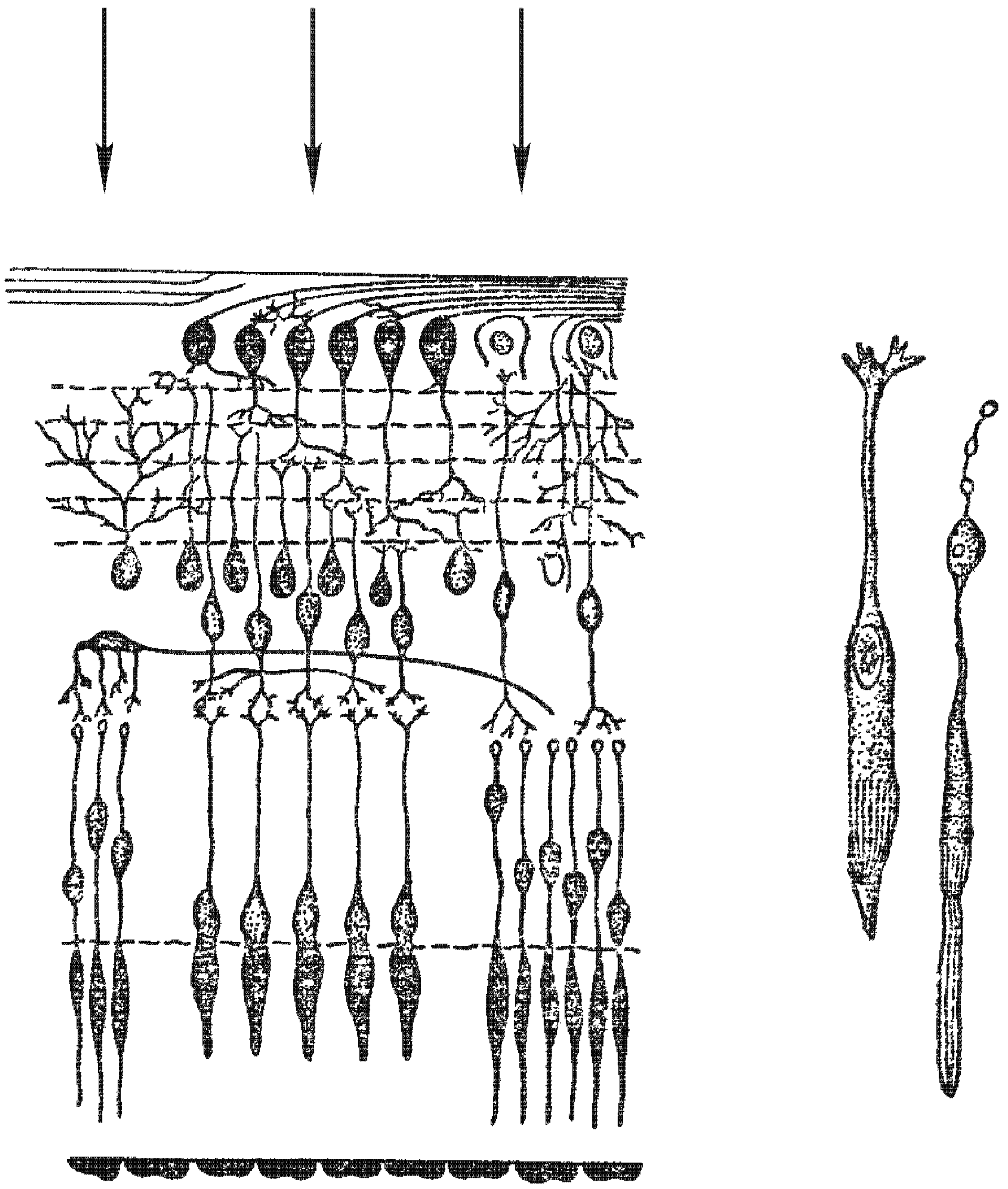
Схематический нормальный глаз по Гульстранду

(Все длины — в миллиметрах, все расстояния — от вершины роговой оболочки)

	При расслабленной глазной мышце	При максимальном напряжении мышцы
Расстояние до задней поверхности роговицы	0,5	
Расстояние до передней поверхности хрусталика	3, 6	3, 2
Расстояние до задней поверхности хрусталика	7,2	
Радиус передней поверхности роговицы	7,7	
Радиус задней поверхности роговицы	6,8	
Радиус передней поверхности хрусталика	10	5, 33
Радиус задней поверхности хрусталика	−6	−5,33
Оптическая сила глаза в диоптриях	58,64	70,57
Расстояние до передней главной плоскости	1,348	1,772
Расстояние до задней главной плоскости	1,602	2,086
Расстояние до переднего фокуса	−15,707	−12,397
Расстояние до заднего фокуса	24,387	21,016
Переднее фокусное расстояние	17,055	14,169
Заднее фокусное расстояние	−22,785	−18,930
Расстояние до центральной ямки	24	
Расстояние до входного зрачка	3,047	2,668
Расстояние до выходного зрачка	3,667	3,212
Увеличение в зрачках	0,909	0,941
Расстояние до ближней точки		102,3

га радиуса ≈ 0,2 мм, палочек нет, есть одни только колбочки. Число их здесь достигает 180 тысяч на квадратном миллиметре, так что среднее расстояние между центрами соседних колбочек ≈ 0,0025 мм.

В том месте, где внутрь глаза входит глазной нерв, т.е. вокруг точки s на рис. 79, на сетчатке нет ни палочек, ни колбочек. Это место называется *слепым пятном*. Его легко обнаружить, закрыв левый глаз и смотря правым на крестик рис. 81. Если глаз приближать к рисунку, то на расстоянии 15–17 см кружок становится невидимым. Это получается тогда, когда изображение кружка попадает на слепое пятно.



а *б*

Рис. 80

Палочки обладают значительно большей чувствительностью к свету, чем колбочки. Ночью и в сумерки (при освещенности меньше 0,01 лк) зрительное ощущение вызывается за счет раздражения одних только палочек. Зато колбочки способны *различать цвета*, а палочки этой способностью не обладают. Практически цветное зрение начинается с освещенностей ≈1 лк. При освещенностях выше 10^2 – 10^3 лк зрение осуществляется почти исключительно колбочками.

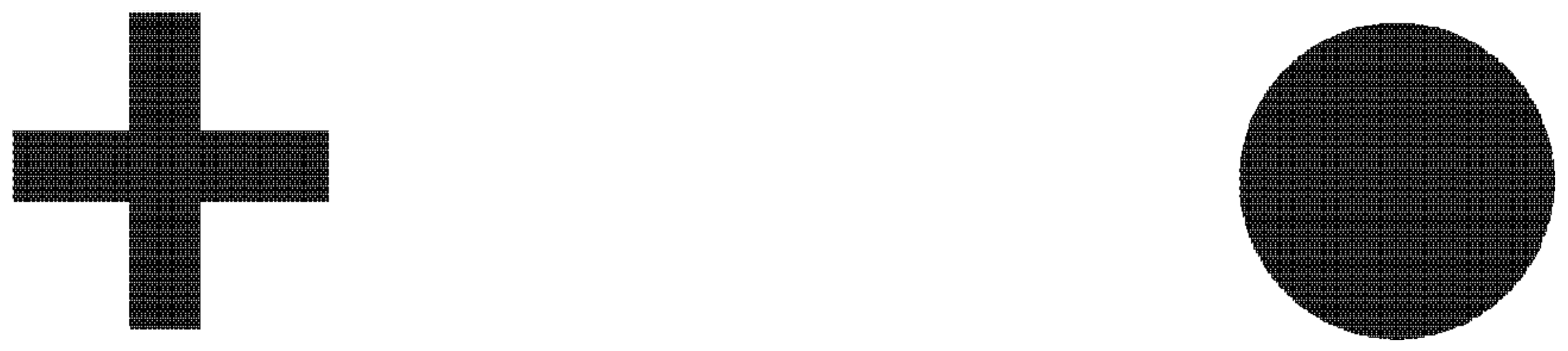


Рис. 81

Детали предмета различаются наиболее отчетливо, когда его изображение получается на желтом пятне и в особенности на центральной ямке сетчатки. Это — так называемое *фовеальное зрение*, в отличие от *периферического зрения*, при котором изображение предмета получается вдали от центральной ямки. При периферическом зрении человек видит предметы значительно менее четко, различая в них мало деталей. Благодаря наличию на периферии в небольшом количестве колбочек, при периферическом зрении ощущение цветов имеет место, но только при условии, что яркость света достаточно велика, чтобы его воспринимали колбочки.

7. Две близкие светящиеся точки глаз воспринимает раздельно, а не как одну слившуюся точку только при условии, что угловое расстояние между этими точками не меньше определенного предела. Этот предел называется *разрешаемым угловым расстоянием*, а его обратная величина — *разрешающей способностью* или *остротой зрения глаза.* За единицу остроты зрения врачи принимают остроту зрения такого глаза, который разрешает угол в одну угловую минуту. Это — в среднем острота зрения нормального глаза.

Острота зрения максимальна, когда изображение приходится на центральную ямку сетчатки. Здесь плотность расположения колбочек максимальна, а каждая колбочка соединена с отдельным волокном зрительного нерва (число этих волокон — несколько миллионов). Мозг реагирует на раздражение каждой колбочки. Когда изображения двух близких светящихся точек приходятся на одну и ту же колбочку, эти точки действуют как одна светящаяся точка. В этом случае разрешения не получается. Для разрешения необходимо, чтобы изображения этих двух точек приходились на *разные колбочки.* На периферии плотность распределения светочувствительных рецепторов — колбочек и палочек — меньше, а каждое волокно зрительного нерва соединено с большим числом этих рецепторов. Вот почему максимальная разрешающая способность глаза получается при фовеальном зрении, когда изображение получается на центральной ямке сетчатки. Зрение же при помощи палочек предназначено не для повышения разрешающей способности глаза, а для *увеличения его чувствительности* в условиях работы при слабой освещенности. Для этого выгодно, как это и есть на самом деле, чтобы с каждым волокном зрительного нерва была соединена не одна палочка, а *большая группа* их. От этого усилится сигнал, передаваемый по этому волокну.

С изложенной точки зрения разрешаемое угловое расстояние глаза $\delta\vartheta$ равно углу, под которым видно из задней узловой точки глаза среднее расстояние δx между двумя соседними колбочками на центральной ямке, т.е. $\delta\vartheta = \delta x/f$, где f — переднее фокусное расстояние. Подставив числа, приведенные выше, получим $\delta\vartheta \approx 30''$. И действительно, наблюдения показали, что для нормального глаза $\delta\vartheta$ несколько меньше одной угловой минуты. Приведенная оценка совпадает с той, какая получается на основе *дифракционных соображений* (см. § 56).

Поле зрения, соответствующее желтому пятну, невелико. На это пятно одновременно может проектироваться картина с угловыми размерами около 6° по горизонтальному направлению и около 4° по вертикальному. Поле зрения центральной ямки еще меньше — около 1° по горизонтали и вертикали. Такая ограниченность поля ясного зрения компенсируется, однако, тем, что живой глаз обладает способностью быстро поворачиваться в глазной впадине, за очень короткое время обегая все точки видимой поверхности рассматриваемого большого предмета. Благодаря этому поле ясного зрения глаза расширяется приблизительно до 150° по горизонтали и 120° по вертикали. Вместе с тем отмеченная подвижность глаза позволяет быстро концентрировать внимание на наиболее важных деталях предмета.

8. Глазу присущи все аберрации обычных оптических систем: и геометрические, и хроматические, и дифракционные. Однако геометрические и хроматические аберрации очень мало заметны, так как глаз использует особые средства снижения их вредного действия. Теперь выяснено, что колбочки имеют форму волноводов. Устройство каждого волновода таково, что лучи, проходящие через периферийную зону зрачка, воздействуют на колбочку слабее центральных лучей. В глазу используются асферические преломляющие поверхности, а хрусталик представляет собой линзу, показатель преломления которой непрерывно возрастает к центру. Это приводит к концентрации света преимущественно вблизи центров кружков рассеяния. Поэтому при не очень ярком освещении края кружков рассеяния мало интенсивны. При ярком же освещении зрачок глаза сильно сужен, а от этого кружок рассеяния уменьшается еще сильнее. Все это ослабляет сферическую аберрацию и кому. Астигматизм косых пучков и дисторсия почти незаметны, так как при подвижности глаза изображение каждой точки бессознательно приводится в наиболее выгодное место сетчатки — центральную ямку. Искривление поверхности изображения компенсируется сферической формой сетчатки. Хроматическая аберрация почти незаметна, ввиду чувствительности сетчатки только к сравнительно узкой части спектра.

9. Остановимся теперь схематически на теории *цветного зрения* Юнга–Гельмгольца (1821–1894), которая лучше других теорий согласуется с наблюдаемыми фактами. Она исходит из того экспериментально установленного факта, что ощущение любого цвета можно получить смешением спектрально чистых излучений красного, зеленого и сине-голубого цветов. На этом основании теория предполагает, что в глазу есть только *три* типа светочувствительных приемников. Они отличаются друг от друга областями спектральной чувствительности. Красный свет воздействует преимущественно на приемники первого типа, зеленый — второго, сине-голубой — третьего. Сложением излучений таких трех цветов в различных количествах можно получить любую комбинацию возбуждений всех трех приемников, а это физиологически эквивалентно получению любого цветового ощущения. Еще не установлено, имеются ли приемники всех трех типов в каждой колбочке или существуют три различных типа колбочек.

Встречаются люди (более 1 % среди мужчин и около 0,1 % среди женщин), зрение которых характеризуется отсутствием приемников одного из трех указанных типов. Они называются *дихроматами*. Дихромат не различает цвета всех излучений, которые для людей с нормальным зрением различаются по степени возбуждения приемника, недостающего у дихромата. Еще реже (примерно раз на миллион людей) встречаются *монохроматы*, у которых есть приемники только одного типа. Такие люди совсем не различают цвета.

10. Чувствительность глаза к излучениям различных длин волн характеризуется *кривой видности*. На этой кривой по оси абсцисс откладывается длина волны, а по оси ординат — *видность* V_λ, т.е. величина, обратная энергетической мощности излучений, которые при оценке глазом воспринимаются как одинаково яркие. Визуальное сравнение яркостей излучений далеких друг от друга длин волн затруднительно. Поэтому для построения кривой видности обычно применяют *метод малых ступеней*, т.е. сравнивают попарно по видимой яркости излучения столь близких длин волн, что разница в цвете не затрудняет такое сравнение. Несмотря на субъективность этого метода, воспроизводимость результатов достаточно хороша, а кривые видности для различных людей не сильно отличаются друг от друга.

Кривая видности среднего нормального глаза при дневном зрении, утвержденная Международной осветительной комиссией, приведена на рис. 82. Она имеет максимум в желто-зеленой части спектра при $\lambda = 555$ нм, условно принимаемый за единицу. При сумеречном зрении, когда работает только палочковый аппарат, кривая видности сохраняет свой общий вид, но смещается в сторону коротких волн с максимумом около 510 нм. При этом область максимальной чувствительности сетчатки смещается на 10–20° в сторону от центральной ямки.

Еще в 1825 г. Пуркинье (1787–1869) наблюдал, что излучения различного цвета, воспринимаемые глазом как одинаково яркие, меняют свою видимую яркость не одинаково, если их ослаблять в одно и то же число раз. Яркость излучений с большей длиной волны уменьшается быстрее, чем с более короткой длиной волны. Если ограничить поле зрения, чтобы оно не превосходило 1,5°, и сконцентрировать сравниваемые излучения в пределах центральной ямки, то свет будет восприниматься только колбочками. Исследования показали, что в этом случае явление

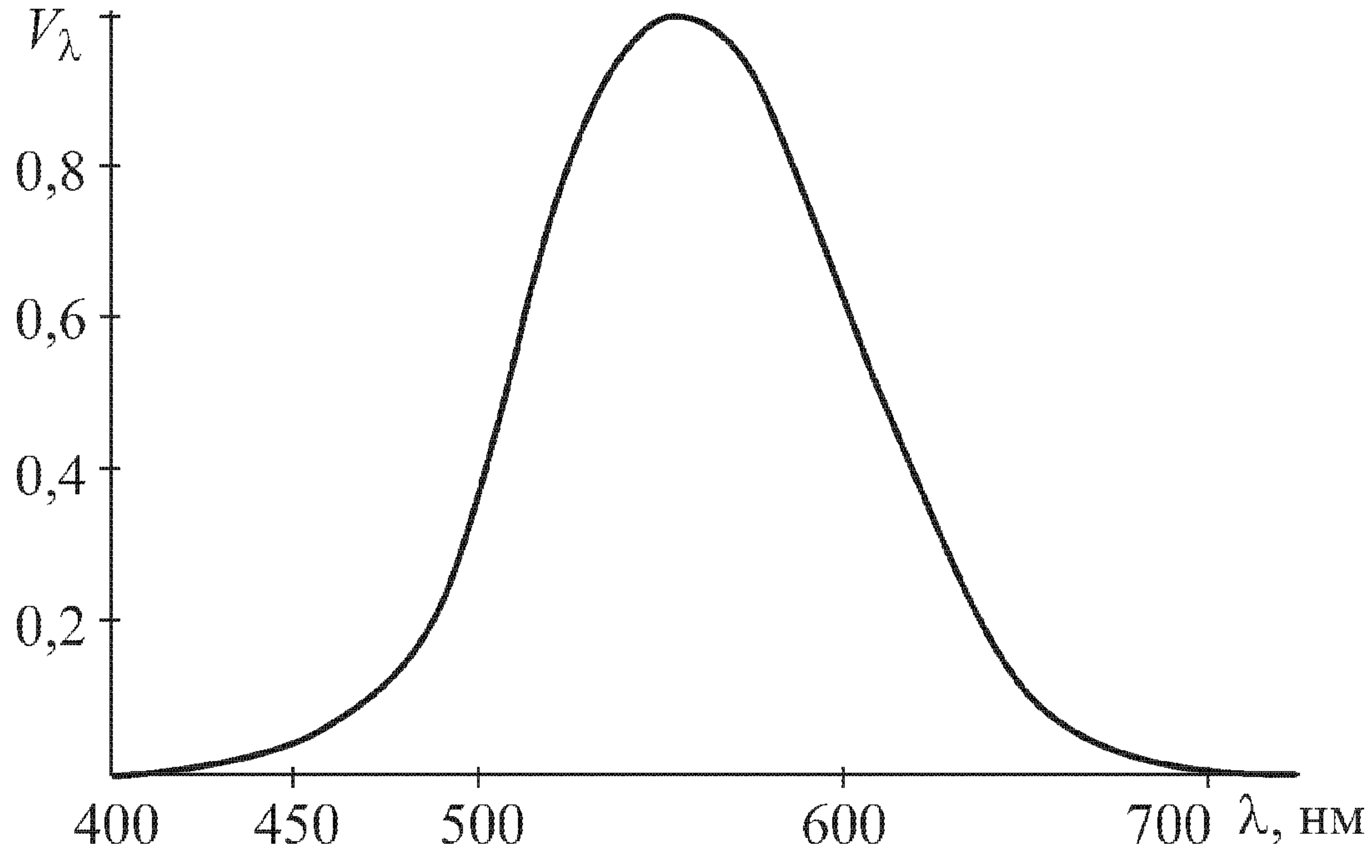


Рис. 82

Пуркинье не наблюдается. Все это хорошо согласуется с двойным механизмом восприятия света: посредством колбочек и посредством палочек, которым соответствуют различные кривые видности.

11. Область, доступная зрительному восприятию глаза, конечно, не обрывается резко на длинах волн 400 и 760 нм. При $\lambda = 400$ нм видность V_λ примерно в 2500 раз, а при $\lambda = 760$ нм — в 20 000 раз меньше, чем в максимуме. В условиях темновой адаптации глаз может видеть в очень слабой степени интенсивные инфракрасные лучи с длинами волн до 950, а ультрафиолетовые — до 300 нм.

Границы видимой области, а также сама форма кривой видности человеческого глаза не случайны. Глаз сформировался в процессе длительной эволюции, приспособившись к условиям освещения земных предметов солнечным светом, а также к условиям сумеречного и ночного освещения. На эту мысль наводит уже то обстоятельство, что на видимую область спектра приходится более 40 % энергии излучения Солнца, хотя она и занимает интервал менее одной октавы. Далее, солнечный спектр вблизи поверхности Земли практически обрывается на длине волны 290 нм. Более короткие волны задерживаются слоем озона в атмосфере. Было бы биологически нецелесообразно, если бы глаз обладал способностью принимать излучение с длинами волн короче 290 нм.

Более того, глаз должен защитить себя от ультрафиолетовых лучей. Они в большинстве случаев химически разлагают органические вещества и могут убивать живые микроорганизмы и клетки. Особенно вредно попадание ультрафиолетовых лучей на сетчатку глаза. Чувствительность сетчатки к ультрафиолетовым лучам довольно велика и, как показал С.И. Вавилов, имеет резкий максимум при $\lambda = 380$ нм. Однако от длинноволновых ультрафиолетовых лучей ($290 < \lambda < 400$ нм), пропускаемых земной атмосферой, глаз защищен собственными средствами. Такие лучи сильно поглощаются внутри глаза, особенно в хрусталике, и лишь ничтожная доля их доходит до сетчатки. С этим и связан сильный спад кривой чувствительности глаза на границе и в области ультрафиолета. При хирургическом удалении хрусталика, если только глаз продолжает функционировать, его чувствительность к ультрафиолетовым лучам заметно повышается[1].

Высокая чувствительность к инфракрасному излучению, если бы глаз обладал таковой, была бы не только нецелесообразна, но и сделала бы невозможной работу глаза в условиях солнечного освещения. Причина этого — в тепловом излучении глаза. При низкой температуре глаза (310 K) все его тепловое излучение приходится на инфракрасную область. Мощность, излучаемую 1 см^2 стенки глаза внутрь глазной полости, нетрудно оценить, пользуясь законом Стефана–Больцмана (см. § 115). Она составляет около 0,105 Вт/см^2. Общая внутренняя поверхность глазной полости ≈ 17 см^2, так что полная мощность излучения глаза внутрь глазной полости будет $\approx 1{,}7$ Вт. Оценим теперь мощность прямого солнечного излучения, попадающего через зрачок глаза, пользуясь значением солнечной постоянной (0,139 Вт/см^2), приведенным в задаче к § 5. Получим $0{,}139\pi r^2 \approx$ $\approx 0{,}017$ Вт (радиус r зрачка глаза принят равным 2 мм). Эта величина примерно в 100 раз меньше мощности собственного теплового излучения глаза. Если бы чувствительность глаза в инфракрасной области спектра была столь же велика, что и в желто-зеленой, то собственное тепловое инфракрасное излучение глаза затмило бы не только рассеянное излучение неба, но и прямой солнечный свет. Работа глаза как органа зрения стала бы абсолютно невозможной. В этом, по мнению С.И. Вавилова (1891–1951), одна из причин, почему нет зрения в инфракрасных лучах. Другая причина заключается в малости энергии инфракрасных квантов света. Процесс зрения должен начинаться либо химическими, либо фотоэлектрическими действиями, а они вызываются отдельными квантами света и не могут происходить, если энергия кванта меньше некоторого минимального предела.

---

[1] Ультрафиолетовые лучи, попадая в глаз, вызывают его флуоресценцию, преобразуясь в синеватый свет, воспринимаемый сетчаткой. Такое косвенное воздействие ультрафиолета проявляется, например, в следующем опыте. Если посмотреть на мощную ртутную кварцевую лампу, излучающую ультрафиолет, через специальное темное стекло, задерживающее видимые, но пропускающее ультрафиолетовые лучи, то благодаря флуоресценции внутри глаза окружающий воздух покажется наполненным синеватым туманом, напоминающим табачный дым.

12. Для возбуждения светочувствительных рецепторов необходимо, чтобы они поглощали свет. Как правило, чувствительность рецептора к свету определенной длины волны должна быть тем выше, чем сильнее поглощает он свет той же длины волны. Это обстоятельство уже давно побудило исследователей искать в сетчатке глаза *светочувствительные пигменты*. Такой пигмент (*родопсин*, или *зрительный пурпур*) был обнаружен в палочках человеческого глаза. Это — вещество розоватого цвета, разлагающееся (выцветающее) на свету и снова восстанавливающееся в темноте. Спектральная кривая поглощения родопсина близка к сумеречной кривой видности. Извлечь светочувствительные пигменты из колбочек сетчатки человека пока не удалось, хотя наблюдения живого глаза через зрачок, по-видимому, подтверждают, что в центральной части сетчатки имеются пигменты, выцветающие на свету.

Глаз человека обладает способностью приспосабливаться к освещенностям, меняющимся в необычайно широких пределах. Прямые солнечные лучи создают освещенности $\sim 10^5$ лк, а в полной темноте глаз способен отличать от темноты предметы с освещенностью 10^{-6} лк. Глаз способен воспринимать световые потоки в интервале 10^{-17}–10^{-5} Вт. Процесс приспособления глаза к тому или иному уровню яркости света называется *адаптацией*. При повышении яркости происходит *световая*, при понижении — *темновая адаптация*. При переходе от яркости ~ 1000 кд/м^2 к темноте чувствительность глаза возрастает в течение часа примерно в 10 миллионов раз. Сначала чувствительность возрастает очень быстро, затем ее рост замедляется и после часа пребывания в темноте уровень чувствительности почти не меняется. Световая адаптация происходит много быстрее. При средних яркостях она продолжается 1–3 минуты. Изменение чувствительности к световому восприятию в столь широких пределах свойственно палочкам. Темновая адаптация колбочек происходит значительно быстрее, причем чувствительность колбочек возрастает всего в 10–100 раз. В состоянии максимальной световой адаптации глаз может без вреда переносить сравнительно большие яркости (например, яркости белых матовых поверхностей, освещаемых прямым солнечным светом). При больших яркостях необходима искусственная защита глаза. Так, наблюдение солнечного затмения можно вести только через закопченное стекло или другой сильно ослабляющий светофильтр. При пребывании на ледниках и в горах на большой высоте необходимо пользоваться темными или цветными очками. Здесь очки необходимы также для защиты от ультрафиолетовых лучей, достигающих на больших высотах значительных интенсивностей и вредно действующих на глаз.

Адаптация глаза происходит автоматически, но механизм ее выяснен не совсем. При резком увеличении яркости света почти мгновенно сужается зрачок. При уменьшении яркости до прежнего уровня зрачок вновь расширяется. Происходит также изменение чувствительности самих рецепторов. Долгое время считалось, что такое изменение чувствительности связано с изменением количества родопсина в палочках. Когда яркость света повышается, родопсин выцветает, и чувствительность глаза понижается. При уменьшении яркости происходит восстановление родопсина и связанное с этим повышение чувствительности. Однако работы последнего времени поставили под сомнение эту точку зрения. Наблюдения показали, что чувствительность глаза к свету меняется всего сильнее, когда уменьшение родопсина еще очень невелико. Наоборот, когда концентрация родопсина резко уменьшается, чувствительность уменьшается незначительно. С изложенной точки зрения это понять трудно. Возможно, играет роль перестройка корковых центров головного мозга, воспринимающих свет, т.е. повышение и понижение их чувствительности.

§ 22. Фотометрические понятия и единицы

1. Излучение в пространстве или в прозрачной однородной среде можно характеризовать *интенсивностью*, *спектральным составом* и *поляризацией*. При этом надо иметь в виду, что пучки строго параллельных лучей являются идеализацией и никогда не встречаются в действительности. Конечной энергией могут обладать лучи, направления которых заполняют *конечные телесные углы*, хотя величина этих углов и может быть очень малой.

Заметив это, возьмем в поле излучения произвольную малую площадку ds. Она должна быть настолько мала, чтобы характеристики излучения практически не изменялись при переходе от одной точки площадки к другой. В то же время линейные размеры площадки ds должны быть велики по сравнению с длинами волн излучения, чтобы к излучению можно было применять понятия и законы геометрической оптики. Проведем через площадку ds лучи, заполняющие какой-то телесный угол Ω.

Энергия, переносимая этими лучами в единицу времени, называется *лучистым потоком* Φ, проходящим через площадку ds в телесный угол Ω. Если телесный угол $d\Omega$ бесконечно мал, а площадка ds перпендикулярна к его оси, то лучистый поток можно представить в виде $d\Phi = I\,ds\,d\Omega$. Величина I есть лучистый поток, отнесенный к единичной площадке, перпендикулярной к излучению, и к единице телесного угла. Она называется *интенсивностью лучистого потока*, или *лучистостью излучения* в направлении оси телесного угла $d\Omega$. Если нормаль к площадке ds образует с направлением излучения угол ϑ, то

$$d\Phi = I\,ds\cos\vartheta\,d\Omega,$$

так как лучистые потоки через площадку ds и ее проекцию $ds_{\perp} = ds\cos\vartheta$, перпендикулярную к излучению, одинаковы. Для краткости проекцию $ds\cos\vartheta$ называют *видимой величиной площадки* ds, если ее рассматривать под углом ϑ к нормали. При выполнении принципа суперпозиции лучи, проходящие в данный момент времени через определенную точку среды, совершенно независимы друг от друга. Поэтому для полной характеристики состояния излучения необходимо указать интенсивность лучистого потока по любому из бесчисленного множества направлений, проходящих через каждую точку пространства. При этом надо учитывать отдельно каждое из двух прямо противоположных направлений, поскольку независимость лучей имеет место и для этих направлений.

Энергия, проходящая за время dt через площадку ds в телесный угол $d\Omega$, представится выражением $I\,ds\,d\Omega\,dt\cos\vartheta$. При этом время dt, хотя оно и входит в виде дифференциала, должно быть все же велико по сравнению с периодами колебаний волн, входящих в излучение. Иначе значение мощности, например, монохроматического излучения, как энергии, отнесенной к единице времени, при малом интервале dt, в течение которого измеряется энергия, зависело бы от фазы колебаний в момент начала измерения. Независимость имела бы место только тогда, когда время dt случайно содержало бы целое число колебаний. Если же dt велико по сравнению с периодами колебаний любых волн, входящих в излучение, то измеренная мощность излучения практически не будет зависеть от выбора dt.

2. *Объемной плотностью лучистой энергии* называется энергия, содержащаяся в единице объема пространства. Чтобы выразить ее через интенсивность излучения I, возьмем сначала малый конус лучей с телесным углом $d\Omega$ при вершине (рис. 83). Через площадку ds, перпендикулярную к излучению, за время dt проходит энергия $I\,ds\,d\Omega\,dt$. Если v — скорость распространения излучения, то за время dt эта энергия распространится на расстояние $v\,dt$ и заполнит объем $ds\,v\,dt$, заштрихованный на рис. 83. (Предполагается, что дисперсии нет, так что скорость распространения энергии совпадает с фазовой скоростью v.) Разделив предыдущее выражение на этот объем, найдем $(I/v)\,d\Omega$ — плотность лучистой энергии, создаваемую рассматриваемым элементарным пучком лучей. Полная плотность лучистой энергии u найдется интегрированием этого выражения по всем телесным углам:

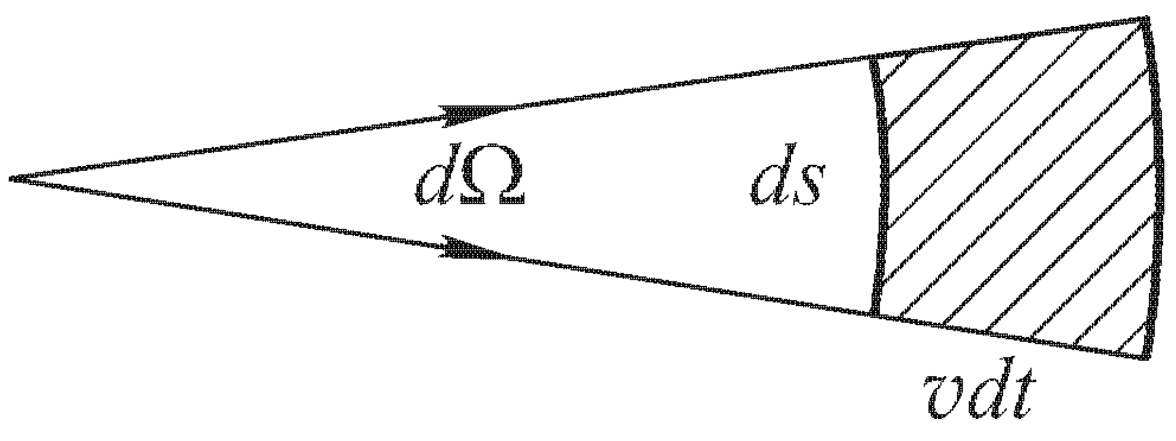


Рис. 83

$$u = \frac{1}{v}\int I\,d\Omega, \tag{22.1}$$

причем величина I может зависеть от направления лучей. Когда I одинакова по всем направлениям, то

$$u = 4\pi I/v. \tag{22.2}$$

3. Величины Φ, I, u можно подвергнуть *спектральному разложению* по частотам или длинам волн. Строго монохроматическое излучение, как и параллельные пучки лучей, никогда не реализуется в действительности. Каждое излучение, обладающее конечной энергией, *занимает конечный интервал частот или длин волн.* Например, величину u можно представить в виде

$$u = \int\limits_0^\infty u_\nu(\nu)\,d\nu = \int\limits_0^\infty u_\omega(\omega)\,d\omega = \int\limits_0^\infty u_\lambda(\lambda)\,d\lambda. \tag{22.3}$$

Подынтегральные выражения $u_\nu\,d\nu$, $u_\omega\,d\omega$, $u_\lambda\,d\lambda$ имеют смысл объемной плотности лучистой энергии в спектральных интервалах $(\nu, \nu+d\nu)$, $(\omega, \omega+d\omega)$, $(\lambda, \lambda+d\lambda)$ соответственно. Ради краткости эти термины обычно применяют к самим функциям $u_\nu(\nu)$, $u_\omega(\omega)$, $u_\lambda(\lambda)$, опуская дифференциалы $d\nu$, $d\omega$, $d\lambda$, хотя это и не совсем логично. При установлении связи между этими функциями надо, конечно, приравнивать друг другу дифференциалы $u_\nu\,d\nu$, $u_\omega\,d\omega$, $u_\lambda\,d\lambda$, поскольку они представляют в различных формах *одну и ту же величину*, если только соответствие между λ, ν и ω устанавливается соотношениями $\lambda = c/\nu = 2\pi c/\omega$. Таким путем находим

$$u_\nu = (\lambda^2/c)u_\lambda. \tag{22.4}$$

Говоря о спектральном распределении, надо указывать, идет ли речь о функции u_ν (распределение по частотам) или о функции u_λ (распределение по длинам волн). Так, в спектре излучения Солнца функция u_ν имеет максимум в инфракрасной области приблизительно при $\lambda = 880$ нм, а функция u_λ — в желто-зеленой части приблизительно при $\lambda = 500$ нм.

Далее, интенсивность излучения I в каждой точке пространства можно представить в виде суммы интенсивностей двух излучений, линейно поляризованных во взаимно перпендикулярных плоскостях. Однако для наших целей в этом пока нет необходимости.

4. Все приведенные энергетические характеристики излучения измеряются в *механических единицах*, например по производимому ими тепловому действию. Так, в системе СИ лучистый поток измеряется в *ваттах* (Вт), интенсивность излучения — в *ваттах на стерадиан-квадратный метр* (Вт/ср·м^2), объемная плотность лучистой энергии — в *джоулях на кубический метр* (Дж/м^3). Такие единицы применяются, например, в теории теплового излучения. Однако в видимой области спектра представляет интерес характеризовать излучение по *зрительному* или *световому ощущению*, оцениваемому по действию света на глаз человека. Соответствующие характеристики и их единицы называются *световыми*, или *фотометрическими*, в отличие от *энергетических* величин и единиц, о которых говорилось выше.

Световые измерения, конечно, содержат немалый элемент субъективизма, поскольку световые впечатления, получаемые различными людьми от одного и того же светового потока, не совсем совпадают между собой. Когда указывают количественные соотношения световых единиц с энергетическими, то при этом имеют в виду не индивидуальный, а какой-то *средний человеческий глаз*. Сами названия величин, характеризующих излучение, при световых измерениях несколько изменяют. Вместо лучистой энергии говорят о *световой энергии*, вместо лучистого потока — о *световом потоке*, вместо интенсивности излучения — об *интенсивности света* и т.п. Вообще, вместо энергетических величин и единиц вводят *световые* или *фотометрические* величины и единицы. Мы не будем вводить для энергетических и фотометрических величин различные обозначения, так как в каждом отдельном случае будет ясно, в каком смысле следует понимать ту или иную величину.

5. *Силой света источника* $\mathcal{J}$ в заданном направлении называют световой поток, посылаемый им в этом направлении и отнесенный к единице телесного угла. (Когда пользуются энергетическими единицами, то говорят об *энергетической силе источника*, измеряя эту величину в Вт/ср.) Обычно это понятие относят к *точечному источнику света*, т.е. характеризуют им источник на расстояниях, больших по сравнению с его линейными размерами. Вообще говоря, величина $\mathcal{J}$ зависит от направления излучения. Световой поток, посылаемый точечным источником в телесный угол $d\Omega$, определяется выражением $d\Phi = \mathcal{J}\,d\Omega$.

Полный световой поток, исходящий от источника,

$$\Phi = \int \mathcal{J}\, d\Omega,$$

причем интеграл берется по полному телесному углу 4π. Величина $\mathcal{J}_0 = \Phi/(4\pi)$ называется *средней сферической силой света источника.*

Единицей силы света источника в системе СИ служит *кандела* (старое название — *свеча*). Это — *основная фотометрическая единица.* Она осуществляется с помощью светового эталона в виде абсолютно черного тела при температуре затвердевания чистой платины (2046,6 К) и давлении 101 325 Па. Устройство эталона показано на рис. 84 (*1* — платина, *2* — трубочка из плавленой окиси тория, *3* — сосуд из плавленой окиси тория, *4* — засыпка из окиси тория, *5* — сосуд из кварца). Нагрев и расплавление платины производятся токами высокой частоты. Излучателем света служит трубочка *2*, стенки которой по всей длине имеют одну и ту же температуру, равную температуре окружающей нагретой платины. *Кандела* (кд) *есть сила света, излучаемого в направлении нормали с* 1/60 см^2 *излучающей поверхности указанного светового эталона.* До введения этого эталона (он был введен с 1 января 1948 г.) единицей силы света источника служила *международная свеча* (м. св.), равная 1,005 кд. Она осуществлялась специально изготовленными электрическими лампочками.

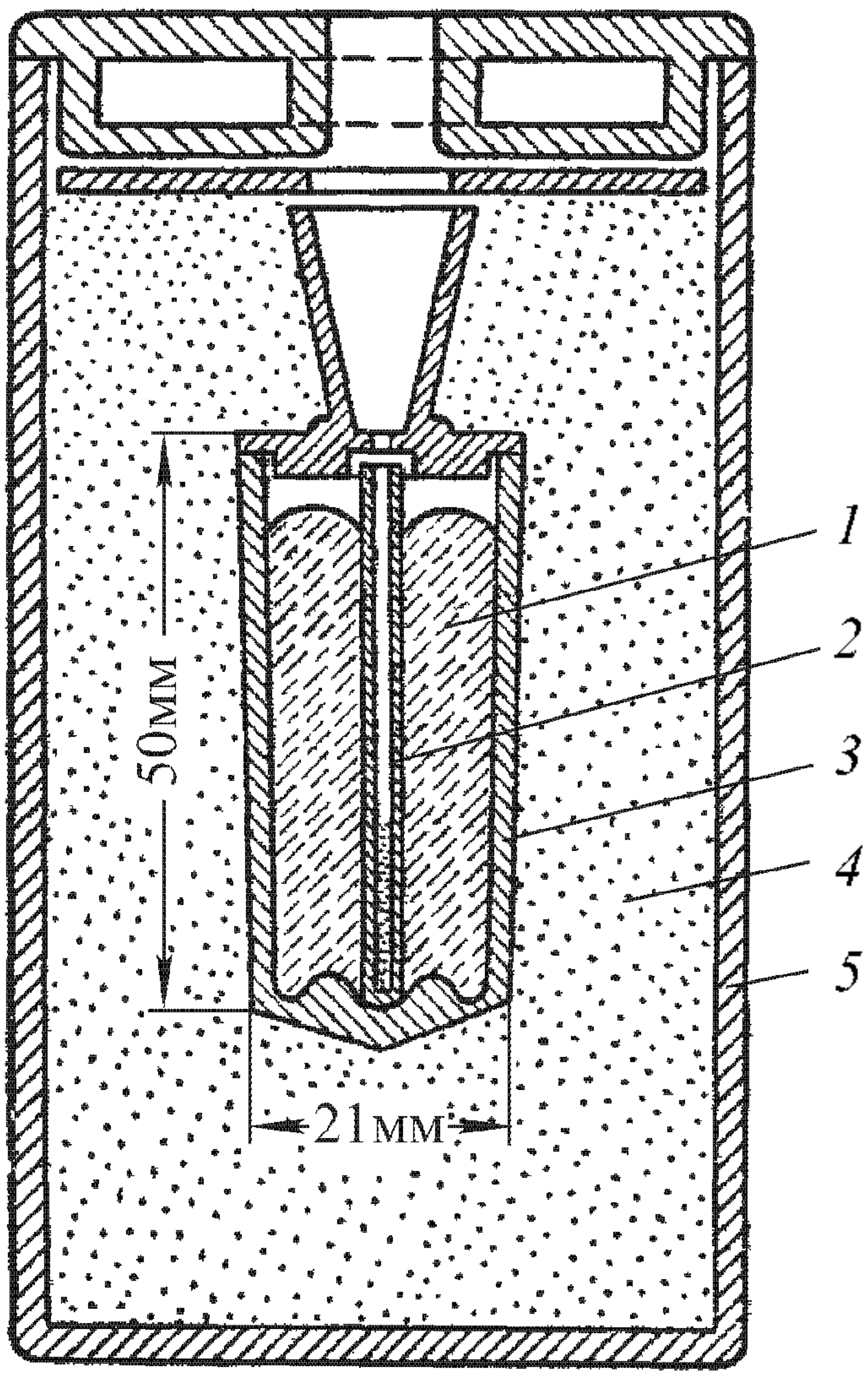


Рис. 84

Единица светового потока есть *люмен* (лм) — световой поток, посылаемый источником в 1 канделу внутрь телесного угла в 1 стерадиан.

Лучистый поток, измеренный в ваттах, можно рассматривать и как световой поток, измеряя его в люменах. *Световая эффективность лучистого потока*, или просто *световая эффективность*, есть число люменов, соответствующее мощности в один ватт (лм/Вт). Обратная величина (Вт/лм) называется *механическим эквивалентом света.* Из-за различной чувствительности глаза к различным участкам спектра обе эти величины зависят от длины волны λ. Принято приводить их значения для $\lambda = 555$ нм, где чувствительность глаза максимальна. Для такой длины волны эти величины по новейшим измерениям равны соответственно:

$$625 \text{ лм/Вт и } 0{,}00160 \text{ Вт/лм}.$$

Пользуясь кривой видности среднего человеческого глаза, приведенной на рис. 82, нетрудно найти значения этих величин для любой длины волны видимого спектра излучения.

6. Из закона сохранения энергии следует, что полный световой поток, посылаемый источником, не может быть увеличен никакими отражающими и преломляющими устройствами, по крайней мере пока они остаются неподвижными. Такие устройства могут только *перераспределять световые потоки* по различным направлениям, что и осуществляется, например, прожекторами. Если точечный источник света помещен в прозрачной однородной среде, то на любых расстояниях от него остается постоянным не только полный поток Φ, испущенный источником в какой-либо момент времени, но и световой поток $d\Phi = \mathcal{J}\, d\Omega$ в пределах любого телесного угла $d\Omega$, исходящего из источника. Ввиду того, что телесный угол $d\Omega$ никак не связан с расстоянием r до источника, не будет зависеть от r и сила света источника $\mathcal{J}$. Интенсивность света I на расстоянии r найдется делением $d\Phi$ на площадь $ds = r^2\, d\Omega$ перпендикулярного сечения рассматриваемого элементарного пучка лучей. Это дает

$$I = \mathcal{J}/r^2, \tag{22.5}$$

т.е. *интенсивность света обратно пропорциональна квадрату расстояния до точечного источника.*

7. Световой поток, приходящийся на единицу площади освещаемой поверхности, называется *освещенностью* E этой поверхности. (При энергетических измерениях вместо этого термина пользуются термином *энергетическая освещенность*, или *облученность*, единицей которой служит Вт/м^2.) Пусть источник точечный, а лучи падают под углом ϑ к нормали к освещаемой поверхности. Тогда $d\Phi = \mathcal{J}\, d\Omega = \mathcal{J}\, ds \cos\vartheta / r^2$. Разделив на площадь поверхности ds, получим

$$E = \frac{\mathcal{J}}{r^2} \cos\vartheta. \tag{22.6}$$

Таким образом, *освещенность, создаваемая точечным источником в отсутствие поглощения, обратно пропорциональна квадрату расстояния до него и прямо пропорциональна косинусу угла между направлением падающих лучей и нормалью к* *освещаемой поверхности.* Первая часть этого утверждения, а также формула (22.5) называются *законом обратных квадратов.* Единица освещенности есть *люкс* (лк) — освещенность, создаваемая световым потоком в 1 люмен, равномерно распределенным по площади в 1 м^2. Освещенность в 1 лк создается точечным источником силою в 1 кд на внутренней поверхности шара радиуса 1 м, если он помещен в центре этого шара и излучает равномерно по всем направлениям.

Чтобы составить конкретное представление о величине люкса, приведем некоторые цифры. Освещенность от Солнца вне земной атмосферы на среднем расстоянии Земли от Солнца $1{,}35\cdot 10^5$ лк. Освещенность в одну-две десятых люкса создает ночью при безоблачном небе полная Луна. Освещенность, создаваемая молодой Луной или Луной на ущербе, порядка нескольких сотых люкса. Безоблачное звездное небо создает ночью освещенность в тысячные доли люкса. Освещенность в десятитысячные доли люкса позволяет с трудом ориентироваться ночью. При освещенности порядка одного люкса можно с трудом читать. Скорость чтения быстро нарастает при увеличении освещенности до 50 лк. При дальнейшем увеличении освещенности до 100–150 лк она растет более медленно, а дальше этого предела возрастание скорости чтения становится малоощутимым. Освещенность 50 лк уже удовлетворительна для чтения и письма. Инструкциями по охране труда установлены определенные нормы минимальной освещенности рабочих помещений. Освещенность рабочей поверхности (стола) ни для каких видов работ не должна быть меньше 10 лк. При очень тонкой работе, связанной с различением мелких деталей, черточек, букв, рисунков (угол зрения меньше $2'$), требуется освещенность не менее 200 лк. В классах и аудиториях на столах учащихся и черных досках освещенность должна быть не менее 75 лк.

8. Для протяженных (не точечных) источников света вводится понятие *поверхностной яркости*, или просто *яркости* B. (При энергетических измерениях вместо этого термина употребляют термин *энергетическая яркость* и для нее вводят единицу Вт/(ср·м^2).) Понятие поверхностной яркости неприменимо для точечных источников, т.е. источников, угловые размеры которых лежат за пределами разрешающей способности глаза или оптического инструмента, которым он вооружен. Возьмем на поверхности излучающего источника малую площадку ds и проведем от нее световой луч под углом ϑ к ее нормали. При рассмотрении излучения в этом направлении более существенна не сама площадка ds, а ее *видимая величина*, т.е. $ds\cos\vartheta$ (см. п. 1). Яркостью поверхности в рассматриваемом направлении называется световой поток $d\Phi$, исходящий из площадки ds в этом направлении, отнесенный к единице телесного угла и к единице ее видимой величины:

$$B_\vartheta = \frac{d\Phi}{d\Omega\, ds \cos\vartheta} = \frac{d\mathcal{J}}{ds \cos\vartheta}, \tag{22.7}$$

где $d\mathcal{J} = d\Phi/d\Omega$ — сила света площадки ds в том же направлении (рис. 85). Буква B снабжена индексом ϑ, так как яркость, вообще говоря, зависит от угла ϑ, под которым рассматривается площадка ds.

Источники света, поверхностная яркость которых B не зависит от направления излучения, называются источниками, подчиняющимися *закону Ламберта* (1728–1777). Для таких источников, как видно из (22.7), сила света $d\mathcal{J}$ элементарной площадки ds пропорциональна $\cos\vartheta$. Однородный светящийся шар, подчиняющийся закону Ламберта, кажется одинаково ярким в середине и по краям. Такие наблюдения и привели Ламберта к формулировке своего закона. В действительности от закона Ламберта наблюдаются большие отступления.

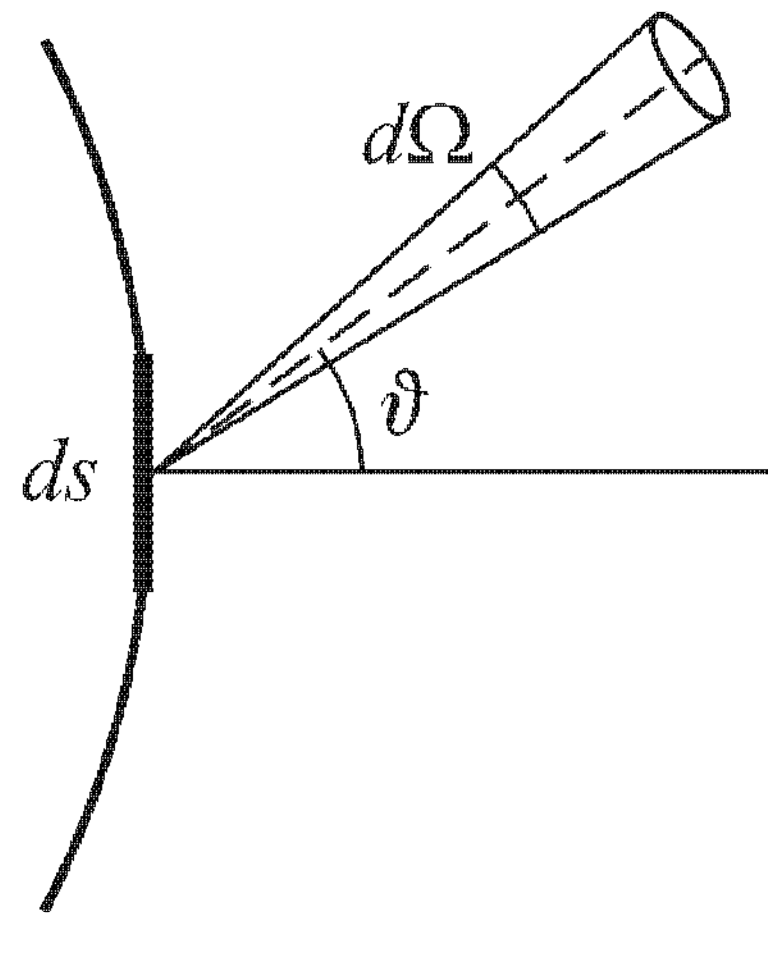


Рис. 85

В § 113 будет показано, что при температурном излучении поверхность непрозрачного тела излучала бы по закону Ламберта, если бы коэффициент отражения света от этой поверхности для каждой длины волны не зависел от угла падения. Для гладких поверхностей, отражающих зеркально, это условие не выполняется (см. § 65). Но для матовых поверхностей, отражающих диффузно, оно может выполняться с той или иной степенью приближения. Для таких поверхностей при температурном излучении приближенно соблюдается закон Ламберта. Он строго справедлив при температурном излучении абсолютно черного тела. Матовые поверхности, например освещенная белая поверхность тела, покрытая окисью магния, или наружная поверхность колпака из хорошего молочного стекла, освещенного изнутри, являются источниками, довольно хорошо подчиняющимися закону Ламберта. Однако к этим случаям вывод закона Ламберта, приводимый в § 113, неприменим, так как в них речь идет не о самосветящихся телах и температурном излучении, а о телах, рассеивающих свет от посторонних источников.

Единицей яркости является *кандела на квадратный метр* (кд/м^2). Это — яркость плоской поверхности, сила света которой в перпендикулярном направлении составляет одну канделу с каждого квадратного метра. Если при тех же условиях сила света равна одной канделе с каждого квадратного сантиметра, то соответствующая единица называется *стильб* (сб). Очевидно, 1 сб = 1 кд/см^2 = 10^4 кд/м^2. В табл. 4 приведены значения яркости некоторых светящихся поверхностей.

Таблица 4

Источник света	Яркость, кд/м^2	Источник света	Яркость, кд/м^2
Ночное безлунное небо	Около $1\cdot 10^{-4}$	Кратер обычной электрической угольной дуги	$1{,}5\cdot 10^{8}$
Полная Луна, видимая сквозь атмосферу	$1\cdot 10^{3}$	Шаровая ртутная лампа сверхвысокого давления (СВДШ)	$1{,}2\cdot 10^{9}$
Пламя обычной стеариновой свечи	$2{,}5\cdot 10^{3}$	Солнце	$1{,}5\cdot 10^{9}$
Ясное дневное небо	$1{,}5\cdot 10^{4}$	Импульсная стробоскопическая лампа (ИСШ)	$1\cdot 10^{11}$
Газосветная лампа	$5\cdot 10^{4}$		
Спираль газонаполненной лампы накаливания	$5\cdot 10^{6}$		

9. *Светимостью* K называется полный световой поток, посылаемый единицей светящейся поверхности в одну сторону, т.е. в телесный угол $\Omega = 2\pi$. Ее единица такая же, что и единица освещенности, т.е. лм/м^2. (При энергетическом рассмотрении вместо этого термина употребляется *энергетическая светимость*, или *излучательность*, а единицей служит Вт/м^2.) Так как световой поток с единицы поверхности в телесный угол $d\Omega$ равен $d\Phi = B_\vartheta \cos\vartheta\, d\Omega$, то

$$K = \int B_\vartheta \cos\vartheta\, d\Omega = 2\pi \int\limits_0^{\pi/2} B_\vartheta \cos\vartheta \sin\vartheta\, d\vartheta. \tag{22.8}$$

Для поверхностей, излучающих по закону Ламберта, яркость $B_\vartheta = B$ не зависит от угла ϑ, а потому в этом случае

$$K = \pi B. \tag{22.9}$$

10. Рассчитаем теперь освещенность, создаваемую протяженным источником света с известной поверхностной яркостью B_υ. Возьмем на поверхности источника бесконечно малую площадку ds (рис. 86). Пусть ds' — освещаемая площадка, r — расстояние между площадками, ϑ и ϑ' — углы, составляемые нормалями к площадкам с прямой MN, соединяющей их. Телесный угол, под которым из площадки ds видна площадка ds', равен $ds' \cos\vartheta'/r^2$. В него площадка ds посылает световой поток $d\Phi = B_\vartheta\, ds \cos\vartheta\, ds' \cos\vartheta'/r^2$, или $d\Phi = B_\vartheta\, d\Omega \cos\vartheta'\, ds'$, где $d\Omega = ds \cos\vartheta/r^2$ — телесный угол, под которым из освещаемой площадки ds' видна излучающая площадка ds. Разделив на ds', найдем освещенность dE площадки ds', создаваемую потоком $d\Phi$.

Полная освещенность E найдется интегрированием полученного выражения по всей видимой поверхности источника:

$$E = \int B_\vartheta \cos\vartheta'\, d\Omega. \tag{22.10}$$

В качестве примера возьмем источник в виде равномерно светящегося диска с поверхностной яркостью $B_\vartheta = B$, не зависящей от угла ϑ. Освещаемую площадку ds' расположим на геометрической оси диска перпендикулярно к ней (рис. 87). Полагая в (22.10) $d\Omega = 2\pi \sin\vartheta'\, d\vartheta'$ и интегрируя, получим

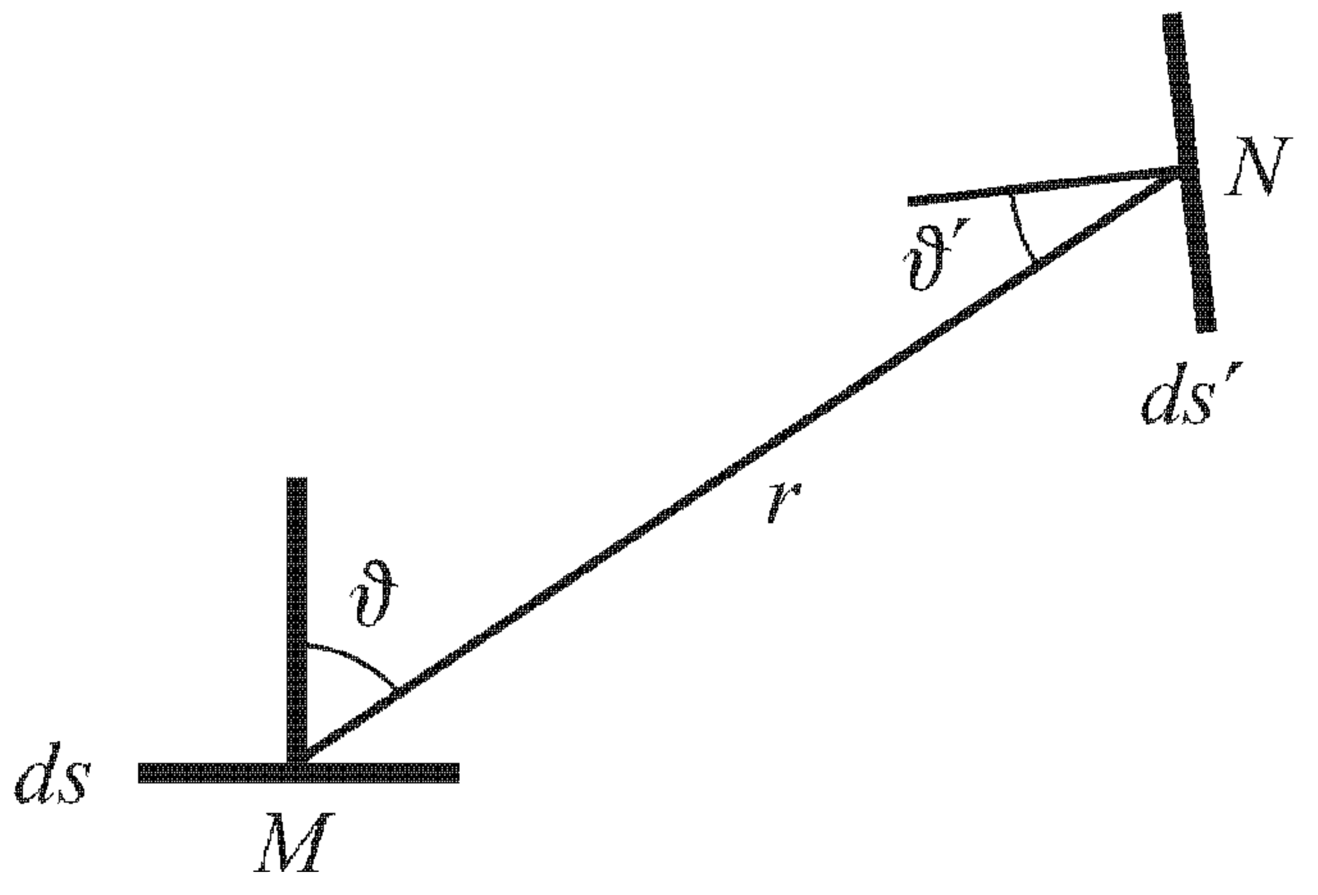


Рис. 86

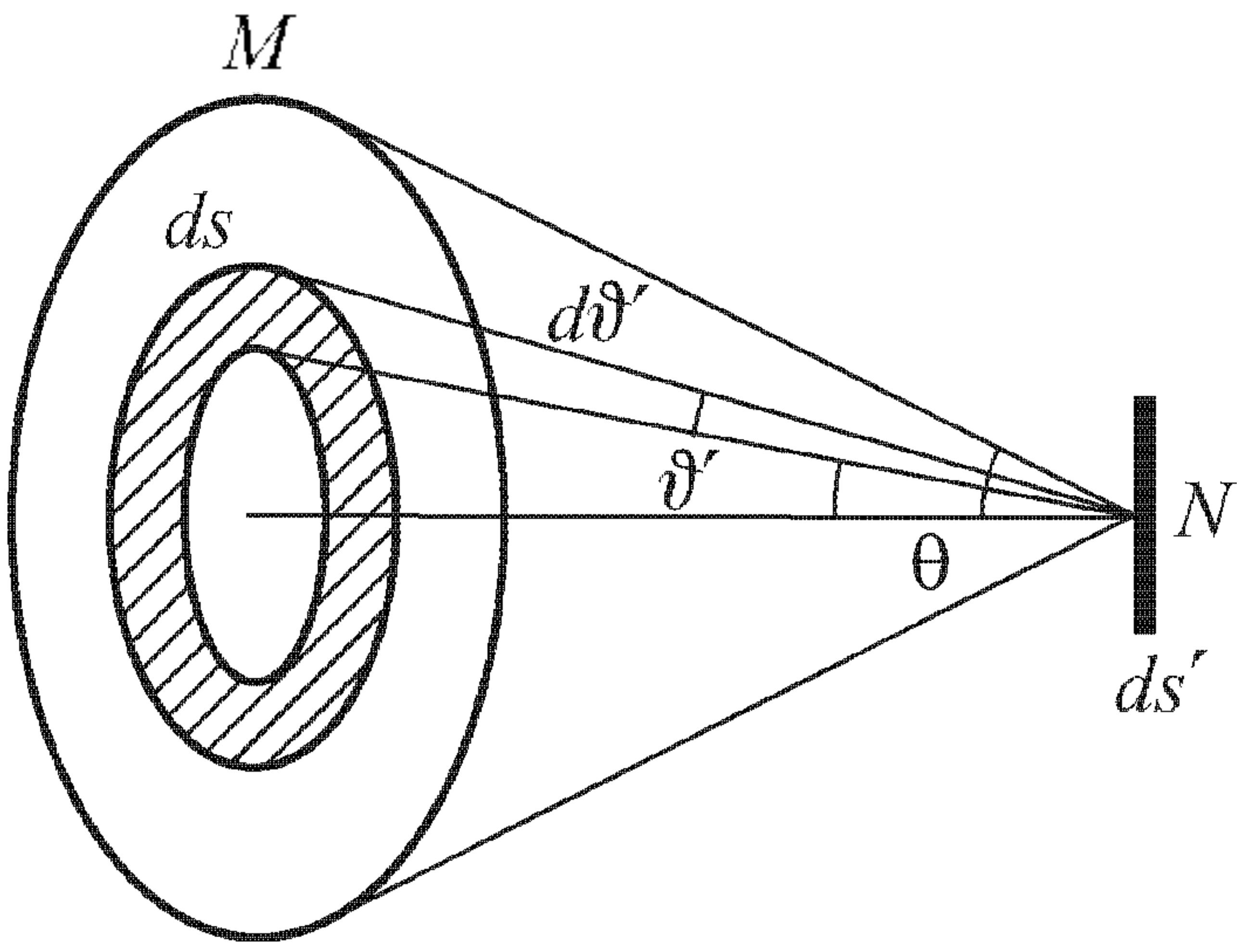


Рис. 87

$$E = \pi B \sin^2\theta, \tag{22.10}$$

где θ — угол, составляемый крайним лучом MN с геометрической осью диска. Формула (22.11) справедлива и для шара, а также для источника произвольной формы, видимая граница которого имеет форму окружности.

11. В нашу задачу не входит изложение методов измерения и описание приборов, применяемых в фотометрии. Ограничимся только схематическим описанием одного из наиболее совершенных фотометров, построенных Луммером (1860–1925) и Бродхуном. Наиболее существенной частью этого фотометра является стеклянный кубик, состоящий из двух прямоугольных стеклянных призм A и B (рис. 88). Гипотенузная плоскость призмы A сошлифовывается под шаровую поверхность, так что от этой плоскости остается только резко очерченный участок в виде плоского круга. Затем призмы гипотенузными плоскостями плотно прикладываются друг к другу (на оптический контакт). Между сравниваемыми источниками света L_1 и L_2 ставится белый экран S, обе поверхности которого одинаково диффузно рассеивают свет. Рассеянный свет отражается двумя одинаковыми зеркалами S_1 и S_2 на кубик фотометра. На призму A попадает свет от источника L_1, на призму B — от источника L_2. Из призмы A свет может попасть в зрительную трубу только через плоский круг на ее гипотенузной плоскости. Наоборот, свет из призмы B может попасть в ту же трубу только после полного отражения от гипотенузной поверхности этой призмы, внешней по отношению к тому же кругу. Труба сфокусирована для наблюдения плоскости указанного круга. Наблюдатель видит два поля зрения различной яркости, граничащие друг с другом вдоль окружности этого круга. Передвигая один из источников L_1 или L_2, добиваются исчезновения видимой границы между обоими полями, чтобы все поле зрения получилось равномерно освещенным. Это произойдет тогда, когда освещенности обеих сторон экрана S станут одинаковыми. Если r_1 и r_2 — расстояния (точечных) источников L_1 и L_2 до экрана в этом положении, то для отношения сил света $\mathcal{J}_1$ и $\mathcal{J}_2$ этих источников по закону обратных квадратов можно написать

$$\frac{\mathcal{J}_1}{\mathcal{J}_2} = \frac{r_1^2}{r_2^2}.$$

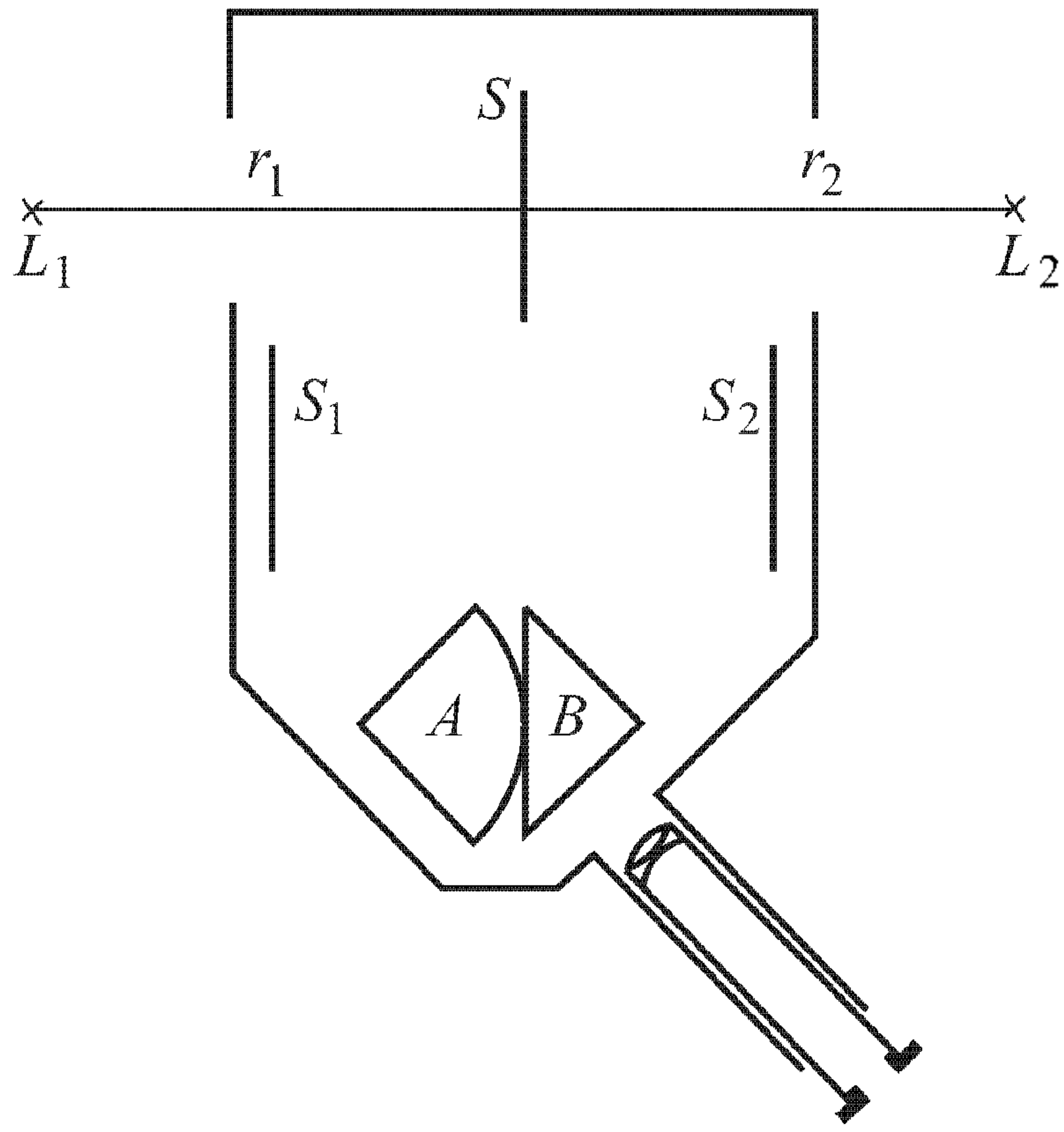


Рис. 88

ЗАДАЧИ

1. Найти освещенность E площадки, создаваемую бесконечной плоскостью, излучающей свет по закону Ламберта с постоянной поверхностной яркостью B. Площадка и плоскость параллельны между собой.

Ответ. $E = \pi B$.

2. Вычислить освещенность E на горизонтальной площадке, освещаемой небесной полусферой, в предположении, что яркость неба B постоянна.

Ответ. $E = \pi B$.

3. Найти освещенность, создаваемую однородным светящимся шаром радиуса a на расстоянии R от его центра, если освещаемая площадка перпендикулярна к радиусу, а поверхность шара излучает по закону Ламберта с поверхностной яркостью B. Показать, что в этих условиях на любых расстояниях от центра шара строго выполняется закон обратных квадратов, т.е. освещенность площадки меняется обратно пропорционально квадрату R.

Ответ. $E = \pi B a^2/R^2$.

4. Найти освещенность, создаваемую равномерно светящимся диском радиуса a на его геометрической оси, перпендикулярной к плоскости диска, если освещаемая поверхность перпендикулярна к этой оси и находится на расстоянии R от центра диска. Диск излучает по закону Ламберта с поверхностной яркостью B. С какой относительной точностью будет выполняться закон обратных квадратов, если диск рассматривать как точечный источник света, помещенный в его центре?

Ответ. $E = \pi B a^2/(R^2 + a^2)$. Если $a/R \ll 1$, то

$$E \approx \frac{\pi B a^2}{R^2}\left(1 - \frac{1}{2}\frac{a^2}{R^2}\right).$$

Закон обратных квадратов выполняется с относительной точностью $a^2/(2R^2)$. При $a/R \sim 1/10$ эта точность будет около 1 %.

5. Предполагая, что излучение Солнца подчиняется закону Ламберта, определить его поверхностную яркость B при наблюдении с поверхности Земли (т.е. с учетом поглощения и рассеяния света земной атмосферой), если освещенность поверхности Земли, создаваемая солнечными лучами при перпендикулярном падении в тех же условиях, — $E \approx 10^5$ лк. Средний телесный угол, под которым солнечный диск виден с Земли, $\Omega = 6{,}8 \cdot 10^{-5}$ ср.

Ответ. $B = E/\Omega \approx 1{,}5 \cdot 10^9$ кд/м^2.

§ 23. Яркость и освещенность оптического изображения. Нормальное увеличение

1. Во всех дальнейших расчетах, если нет специальной оговорки, не будем учитывать потери света при отражении на границах линз, призм и других отражающих поверхностях. Будем также пренебрегать поглощением и рассеянием света. Потери света на отражение довольно значительны. Так, для призматического бинокля они составляют около 50 % (см. § 67). С учетом всех этих потерь яркость и освещенность изображения получились бы несколько меньше вычисленных. Сначала рассмотрим случай, когда размеры предмета заметно превосходят предел разрешения оптической системы. Тогда система дает геометрически подобные изображения, и можно ограничиться точностью геометрической оптики.

Допустим, что предмет помещен на главной оптической оси центрированной системы и имеет форму малой площадки s, перпендикулярной к этой оси (рис. 89 *а*). Изображением площадки s будет какая-то другая площадка s'. Обозначим через B_ϑ яркость площадки s под углом ϑ к ее нормали, или (что то же) к главной оптической оси системы. Соответствующую яркость изображения обозначим через $B'_{\vartheta'}$. Соот-

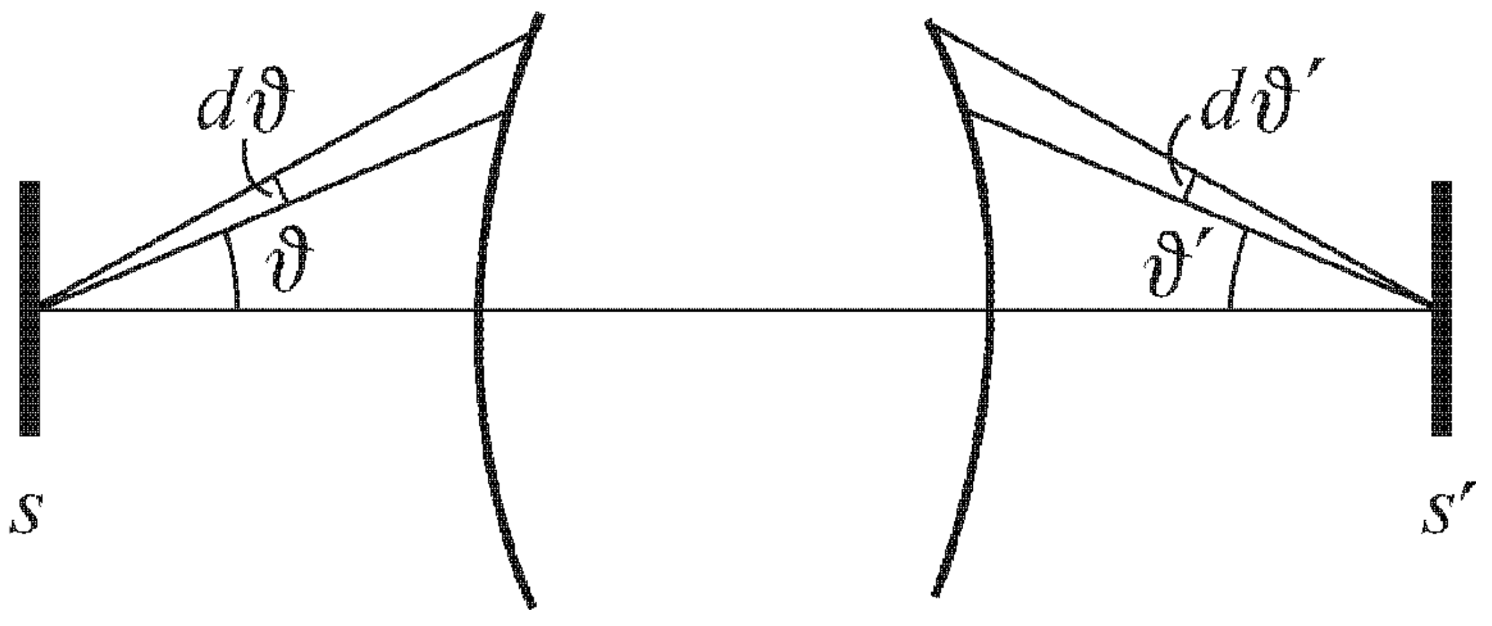


Рис. 89 *а*

ветствие означает, что угол ϑ, составляемый падающим лучом с оптической осью системы, после выхода из нее переходит в ϑ'. Будем предполагать, что в образовании изображения участвуют не только параксиальные, но и широкие пучки лучей, так что должно выполняться условие синусов (18.1). Из него следует

$$sn^2 \sin^2 \vartheta = s'n'^2 \sin^2 \vartheta'. \tag{23.1}$$

Если условие синусов не выполняется, то все результаты, получаемые ниже, также сохраняют силу, но только для параксиальных лучей. Возьмем теперь два конуса с общей вершиной на площадке s, образующие которых составляют с главной оптической осью углы ϑ и $\vartheta + d\vartheta$. (На рис. 89 *а* указаны только сечения этих конусов плоскостью рисунка.) Величина телесного угла, заключенного между этими конусами, будет $d\Omega = 2\pi \sin\vartheta \, d\vartheta$. Световой поток, посылаемый площадкой s внутрь телесного угла $d\Omega$, равен

$$d\Phi = B_\vartheta s \cos\vartheta \, d\Omega = 2\pi B_\vartheta s \sin\vartheta \cos\vartheta \, d\vartheta.$$

В пространстве изображений световой поток $d\Phi$ перейдет в поток

$$d\Phi' = 2\pi B'_{\vartheta'} s' \sin\vartheta' \cos\vartheta' \, d\vartheta',$$

пронизывающий площадку s'. При отсутствии потерь света $d\Phi = d\Phi'$, т.е.

$$B_\vartheta s \sin\vartheta \cos\vartheta \, d\vartheta = B'_{\vartheta'} s' \sin\vartheta' \cos\vartheta' \, d\vartheta'.$$

Учтем теперь соотношение (23.1), дифференцирование которого дает

$$sn^2 \sin\vartheta \cos\vartheta \, d\vartheta = s'n'^2 \sin\vartheta' \cos\vartheta' \, d\vartheta'.$$

В результате получим

$$\frac{B_\vartheta}{n^2} = \frac{B'_{\vartheta'}}{n'^2}, \tag{23.2}$$

чем и определяется яркость изображения. Если показатели преломления пространств предметов и изображений одинаковы ($n = n'$), то $B'_{\vartheta'} = B_\vartheta$, т.е. яркость изображения равна яркости предмета. Как видно из вывода, одинаковость яркостей обусловлена тем, что увеличение площади изображения сопровождается уменьшением в такое же число раз телесного угла, в котором распространяется световой поток. Для источников света, подчиняющихся закону Ламберта, B_ϑ не зависит от угла ϑ. В этом случае нет смысла сохранять индексы ϑ и ϑ', так что можно написать

$$B' = (n'/n)^2 B. \tag{23.3}$$

2. Иначе ведет себя освещенность изображения, проецирующегося на экран. В этом случае величина телесного угла, в котором распространяются лучи после выхода из оптического прибора, не играет роли. Существенны только площади изображения и полный световой поток, концентрирующийся на этой площади. При неизменном световом потоке освещенность будет тем больше, чем меньше площадь изображения.

При количественном исследовании этого вопроса будем предполагать, что изображаемая площадка s имеет форму диска, плоскость которого перпендикулярна к главной оптической оси, и излучает свет по закону Ламберта. Если бы оптического прибора не было, то освещенность экрана, согласно формуле (22.11), была бы равна $E_0 = \pi B \sin^2\theta$, где 2θ — угол зрения, т.е. угол между лучами, исходящими из диаметрально противоположных точек площадки s и направленными к центру площадки s'. При наличии оптического прибора угол схождения крайних лучей с вершиной на площадке s' меняется и становится равным $2\theta'$. Это есть угол между крайними лучами, идущими к центру площадки s' от диаметрально противоположных точек выходного зрачка прибора. Полный световой поток через площадку s', как было показано выше, представляется интегралом

$$\Phi' = 2\pi B' s' \int \sin\vartheta' \cos\vartheta'\, d\vartheta' = \pi B' s' \sin^2\theta',$$

а освещенность площадки s' будет Φ'/s', т.е.

$$E = \pi B' \sin^2\theta' = \pi B \left(\frac{n'}{n}\right)^2 \sin^2\theta'. \tag{23.4}$$

Если $n = n'$, то

$$\frac{E}{E_0} = \left(\frac{\sin\theta'}{\sin\theta}\right)^2. \tag{23.5}$$

Эта формула показывает, что влияние оптического прибора на освещенность изображения сводится к изменению угла между крайними лучами, исходящими из противоположных точек предмета. Например, действие лупы как зажигательного стекла (в отсутствие потерь света) эквивалентно приближению источника (Солнца) на такое расстояние, чтобы он был виден невооруженным глазом под тем же углом зрения, как и входной зрачок (свободное отверстие) лупы из ее главного фокуса.

Если изображение получается в главной фокальной плоскости объектива (как в фотоаппарате), то в параксиальном приближении $\sin\theta' = D'/(2f')$, где D' — диаметр выходного зрачка, а f' — заднее фокусное расстояние объектива. Освещенность изображения будет пропорциональна квадрату этого отношения, а с ним и отношению D/f', где D — диаметр входного зрачка. Отношение D/f' называется *относительным отверстием*, а его квадрат $(D/f')^2$, определяющий освещенность изображения, — *светосилой объектива*.

В иммерсионных микроскопах (см. § 18, п. 4) свет от конденсора, помещенного перед объективом, попадает в пространство, заполненное жидкостью (иммерсией) с показателем преломления n, в которой помещается предмет. Пусть $2\theta_0$ — угол (апертура) между крайними направлениями лучей в конденсоре, которые попадают в микроскоп. Так как сам конденсор находится в воздухе, то по закону преломления $\sin\theta_0 = n\sin\theta$. Световой поток пропорционален $\sin^2\theta_0$, т.е. $(n\sin\theta)^2$. При одной и той же апертуре 2θ световой поток, попадающий в объектив, а с ним и освещенность изображения пропорциональны квадрату показателя преломления n. Величину $n\sin\theta$, следуя Аббе, называют *числовой апертурой*. Таким образом, применение иммерсии увеличивает освещенность изображения. Так как, однако, окончательное изображение получается в воздухе, то и при наличии иммерсии его яркость может быть самое большее равна яркости источника света.

3. Перейдем теперь к зрительной оценке поверхностной яркости предметов, т.е. оценке при наблюдении их глазом.

Отдельные элементы сетчатки — колбочки и группы палочек, соединенные с одним и тем же волокном зрительного нерва, — реагируют на световое раздражение независимо друг от друга. Увеличение освещаемой поверхности сетчатки при сохранении неизменной ее освещенности не усиливает интенсивности светового раздражения отдельного элемента, а увеличивает лишь число освещенных элементов. Если отвлечься от трудно поддающихся учету различных физиологических и психологических факторов, с которыми связаны, например, многочисленные обманы зрения, то следует сказать, что глаз судит о яркости предмета не по общему световому потоку, попадающему в него, а прежде всего по потоку, приходящемуся на единицу площади сетчатки, т.е. по *освещенности* последней. Поэтому, если исключить все субъективные факторы, то вопрос о зрительной оценке яркости сведется к вопросу об освещенности сетчатки.

Освещенность сетчатки невооруженного глаза называется *естественной освещенностью*. Она найдется по формуле (23.4), в которой следует положить $n = 1$, если, как это обычно и бывает, рассматриваемый предмет находится в воздухе. Таким образом,

$$E = \pi B (n' \sin\theta')^2. \tag{23.6}$$

Угол θ' есть угол, под которым из заднего фокуса глаза виден диаметр его выходного зрачка. В формулу (23.6) совсем не входит расстояние до предмета. *При неизменном зрачке глаза освещенность сетчатки E не зависит от этого расстояния, а определяется только поверхностной яркостью предмета.* Вот почему одинаковые фонари, находящиеся на разных расстояниях от нас, кажутся нам одинаково яркими (если, конечно, пренебрежимо малы потери света при распространении его от фонаря до глаза).

Допустим теперь, что глаз вооружен, т.е. наблюдение ведется через зрительную трубу (телескоп) или микроскоп. В этом случае оптический прибор и глаз следует рассматривать как единую оптическую систему, к которой применимы формулы (23.5) и (23.6). Однако величина угла θ' может измениться. Смотреть в оптический прибор следует так, чтобы плоскость выходного зрачка прибора совмещалась с плоскостью входного зрачка глаза. Только тогда поле зрения будет резко очерчено. Угол θ' определяется тем из зрачков, который сильнее диафрагмирует световые лучи. Если входной зрачок глаза меньше выходного зрачка прибора, то угол θ' определяется зрачком глаза. В этом случае получается естественная освещенность сетчатки, т.е. сетчатка будет освещена так же, как и при отсутствии прибора (если, конечно, пренебречь дополнительными потерями света, вносимыми оптическим прибором). Если же входной зрачок глаза больше выходного зрачка прибора, то угол θ' определится зрачком прибора. В этом случае освещенность сетчатки будет меньше естественной. Будем непрерывно увеличивать выходной зрачок прибора, начиная с очень малых размеров. Сначала освещенность сетчатки будет возрастать пропорционально площади выходного (а, следовательно, и входного) зрачка (в параксиальном приближении). Когда выходной зрачок прибора сделается равным входному зрачку глаза, освещенность E достигнет максимума, а затем будет оставаться постоянной.

4. Изложенными соображениями руководствуются при выборе рационального увеличения зрительных труб, телескопов, биноклей и микроскопов. *Нормальным* или *равнозрачковым* увеличением таких приборов называют такое увеличение, когда выходной зрачок прибора равен входному зрачку глаза.

В телескопе, как и во всякой зрительной трубе, апертурной диафрагмой и входным зрачком служит свободное отверстие объектива. Объектив дает изображение в своей задней фокальной плоскости, которая одновременно является передней фокальной плоскостью окуляра. При этом условии изображение в телескопе получается на бесконечности, и его отчетливо будет видеть нормальный глаз в ненапряженном состоянии. Все параллельные пучки лучей после прохождения через телескоп остаются параллельными, т.е. телескоп является телескопической оптической системой. Меняется только ширина пучков. Если падающие лучи параллельны главной оптической оси, то ширина пучка равна диаметру объектива D. Увеличение телескопа N есть отношение угла, под которым виден малый предмет в телескоп, к углу, под которым он был бы виден, если бы рассматривался невооруженным глазом. Как было показано в § 11 (п. 10), для телескопических систем такое увеличение равно отношению ширины падающего пучка параллельных лучей к ширине выходящего пучка. В телескопе ширина выходящего пучка равна диаметру выходного зрачка D'. При нормальном увеличении $D' = d$, где d — диаметр входного зрачка глаза. Таким образом, нормальное увеличение телескопа определяется выражением

$$N_{\text{норм}} = D/d. \tag{23.7}$$

Диаметр d зрачка глаза не остается постоянным, а зависит от яркости. Увеличение, равнозрачковое в дневных условиях, оказывается слишком большим при наблюдении в сумерки и ночью. В последних случаях требуются трубы с большим выходным зрачком, т.е. либо с малым увеличением, либо, что гораздо лучше, с большим объективом.

Освещенность изображения не зависит от увеличения, если оно меньше нормального. Увеличения меньше нормального нецелесообразны, так как при этом используется только центральная часть объектива; периферийные лучи, достаточно удаленные от оптической оси, задерживаются радужной оболочкой и не попадают в зрачок глаза. Нормальное увеличение есть максимальное увеличение, при котором освещенность изображения еще максимальна. При таком увеличении объектив используется полностью. Поэтому окуляры к зрительным трубам подбирают так, чтобы они давали нормальное увеличение.

В некоторых случаях имеет смысл применять увеличения, превосходящие нормальное в 2–4 раза, несмотря на то, что это связано с значительным уменьшением освещенности изображения (соответственно в 4–16 раз). Дело в том, что разрешаемый предел глаза 30–60$''$ соответствует случаю, когда изображение получается на центральной ямке желтого пятна. На периферии разрешаемый предел больше. Он увеличивается также при длительной работе и усталости глаза. В этих условиях выгоднее пойти на некоторое уменьшение освещенности, чтобы получить большие увеличения и возможность различать более мелкие детали предметов. Однако далеко идти в этом направлении нецелесообразно из-за уменьшения освещенности и поля зрения.

При N-кратном увеличении угловые размеры поля зрения на местности уменьшаются в N, а его площадь в N^2 раз. Поэтому в биноклях предпочитают увеличение меньше нормального. Этим достигается большее поле зрения и облегчается нахождение нужного объекта на местности. Бинокли рассчитывают для сравнительно широких выходных зрачков (в пределах 3–5 мм). Так, полевые бинокли с диаметром объектива 3 см имеют 6- или

8-кратные увеличения (диаметр выходного зрачка соответственно 5 и 3,75 мм).

5. Зрительные трубы, как обнаружил еще М.В. Ломоносов (1711–1765), улучшают видимость изображений предметов при слабых освещенностях, например ночью. Конечно, освещенность изображений при этом не только не увеличивается, но даже уменьшается из-за неизбежных потерь света на отражение от поверхностей линз и поглощение в стекле. Однако при слабых освещенностях сильно уменьшается разрешающая способность глаза. Ночью, когда освещенность падает до десятитысячных долей люкса, угол, разрешаемый глазом, возрастает с $\sim 1'$ до $\sim 1^\circ$. Зрительные трубы увеличивают угол зрения, а потому их применение способствует различению контуров и крупных деталей предметов, даже если они неразличимы невооруженным глазом.

В трубах, предназначенных для ночных наблюдений, число линз должно быть минимальным, а сами линзы должны быть простыми, чтобы максимально уменьшить потери света на отражение. Увеличение трубы должно быть как можно ближе к нормальному, чтобы использовать весь световой поток, падающий на объектив. Конечно, нормальное увеличение определяется размерами зрачка глаза в ночных условиях (6–8 мм).

6. Найдем теперь нормальное увеличение микроскопа. Увеличением микроскопа N называется отношение угла φ', под которым виден малый объект в микроскоп, к углу φ, под которым он был бы виден невооруженным глазом, если бы был помещен от него на расстоянии ясного зрения L (см. § 21, п. 3). Если y — линейный размер рассматриваемого объекта, то $\varphi = y/L$. Поместим этот объект $y = AB$ в передней апланатической точке объектива (рис. 89 *б*); для упрощения рисунка сложные объектив и окуляр изображены тонкими линзами, это не отражается на общности выводов и результатов). Его изображение $A'B' = y'$ получится в задней апланатической точке. Так как угол θ' всегда мал, то его синус можно заменить самим углом, записав условие синусов в виде $ny \sin\theta = y'\theta'$. Так же, как и в телескопе, изображение y' должно получаться в передней фокальной плос-

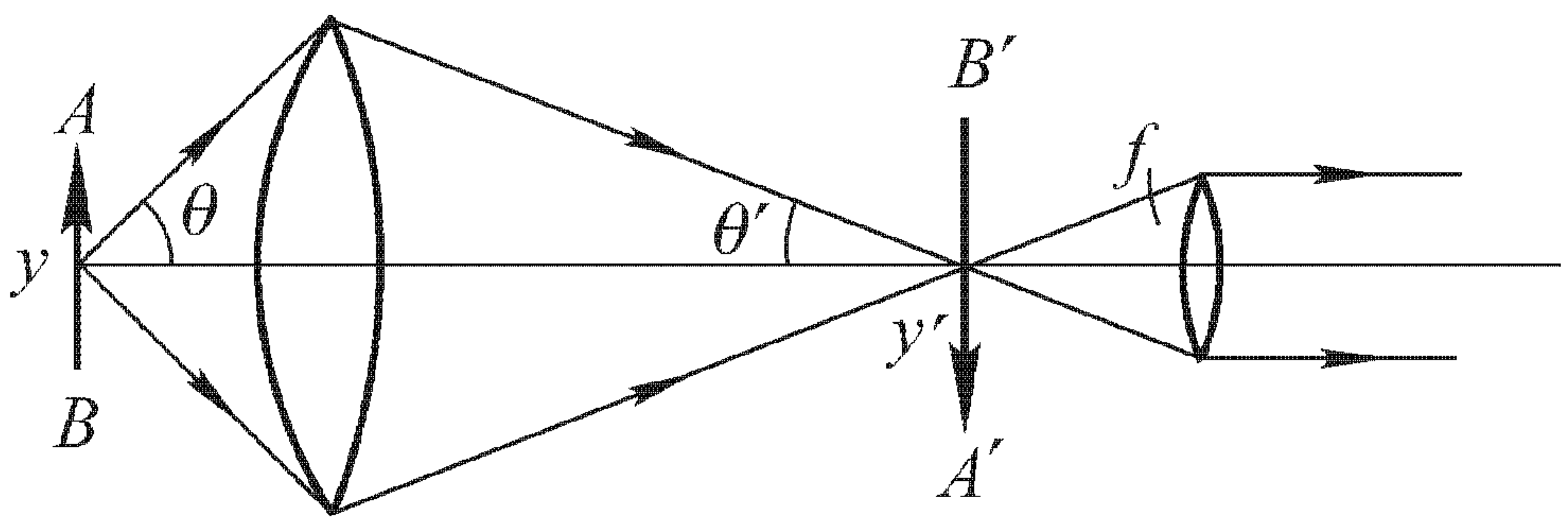


Рис. 89*б*

кости окуляра. Если f — фокусное расстояние окуляра, то $2f\theta'$ будет шириной выходящего пучка лучей. При нормальном увеличении эта ширина должна быть равна диаметру d входного зрачка глаза, т.е.

$$d = 2f\theta' = 2\,\frac{fny\sin\theta}{y'}.$$

Но $y = L\varphi$, $y' = f\varphi'$, так что $y/y' = L\varphi/(f\varphi') = L/(Nf)$. Подставляя это в предыдущее соотношение, получим [1]

$$N_{\text{норм}} = 2Ln\frac{\sin\theta}{d}. \tag{23.8}$$

Нормальное увеличение для микроскопа еще в большей степени есть величина субъективная, чем для телескопа, так как оно зависит от двух параметров глаза: диаметра зрачка d и расстояния ясного зрения L. Характеризуя увеличение микроскопа, обычно полагают $d = 2$ мм, $L = 250$ мм. Для сухих систем ($n = 1$) верхний предел числовой апертуры $n\sin\theta$ составляет 1, для иммерсионных 1,5. Соответствующие нормальные увеличения будут $N = 250$ и $N = 375$. По тем же соображениям, что и для телескопа, от этих значений можно несколько отступать в бóльшую сторону, но не более чем в 2–4 раза. Таким образом, в лучших безиммерсионных микроскопах верхний предел увеличения можно принять равным 1000, а в иммерсионных — 1500. Бóльшие увеличения не только бесполезны, но и вредны. К этому вопросу, а также к вопросу об увеличении телескопа мы вернемся в § 56 и разберем его с новой точки зрения — в связи с разрешающей способностью этих приборов, определяемой волновой природой света.

7. В заключение рассмотрим случай, когда размеры объекта меньше предела разрешения оптической системы. Такой случай типичен при наблюдении звезд невооруженным глазом и в телескоп. В этом случае оптическое изображение не передает формы предмета, а представляет собой дифракционный кружок, окруженный темными и светлыми кольцами (см. рис. 180, с. 316), причем почти весь свет сконцентрирован в центральном кружке. Назначение телескопа при наблюдении звезд состоит не в том, чтобы различать детали рассматриваемого объекта (они находятся за пределами его разрешающей способности), а в том, чтобы *увеличить световой поток*, попадающий в глаз наблюдателя, и тем самым обнаруживать все более и более слабые звезды.

---

[1] Очень часто при выводе формулы (23.8) вводят ненужное предположение, что изображение в микроскопе получается на расстоянии ясного зрения L от глаза. На самом деле это может быть и не так. Справедливость формулы (23.8) никак не связана с тем, на каком расстоянии от глаза получается изображение в микроскопе. Величина L входит в формулу (23.8) совсем по другой причине, как это видно из нашего вывода.

Применять увеличение меньше нормального не имеет смысла, так как при этом будет использоваться не весь объектив. Поэтому в дальнейшем достаточно ограничиться случаем, когда диаметр выходного зрачка телескопа D' меньше или равен диаметру входного зрачка глаза d ($D' \leqslant d$). Тогда размер дифракционного кружка будет определяться величиной D'. Как будет показано в теории дифракции (см. § 55), диаметр дифракционного кружка будет обратно пропорционален D', а его площадь обратно пропорциональна D'^2. Световой поток, попадающий в глаз, $\Phi \sim D^2$, где D — диаметр объектива, а освещенность дифракционного кружка на сетчатке

$$E \sim \frac{D^2}{1/D'^2} = D^2 D'^2.$$

Вводя увеличение телескопа $N = D/D'$, получим

$$E \sim D^4/N^2.$$

Видимость звезды определяется не только освещенностью дифракционного кружка на сетчатке, но и освещенностью окружающего фона. Освещенность фона $E_{\text{фон}}$ определяется яркостью B неба. Ее можно найти по формуле (23.6), в которой под θ' теперь следует понимать угол, под которым из заднего фокуса глаза виден диаметр D' выходного зрачка телескопа. Таким образом, $E_{\text{фон}} \sim BD'^2 = BD^2/N^2$, и следовательно,

$$E/E_{\text{фон}} \sim BD^2. \tag{23.9}$$

Эта формула объясняет, почему в телескоп звезды видны и днем.

ЗАДАЧА

С помощью собирательной линзы можно зажечь папиросу, фокусируя на ней прямой солнечный свет. Можно ли, взяв линзу достаточных размеров (например, объектив большого астрономического телескопа), сделать то же самое, используя свет полной Луны?

Решение. Для простоты будем считать, что излучение Солнца подчиняется закону Ламберта. Прямые солнечные лучи, падающие на Землю перпендикулярно, создают на ней освещенность $E_{\text{C}} = \pi B \sin^2 \vartheta_{\text{C}}$, где B — поверхностная яркость Солнца, а ϑ_{C} — угловой размер его. Из повседневного опыта известно, что такой освещенности для воспламенения папиросы недостаточно. Средняя освещенность поверхности Земли будет в π раз меньше, т.е. $\overline{E}_{\text{C}} = B \sin^2 \vartheta_{\text{C}}$. Если пренебречь поглощением и рассеянием света, то действие линзы сводится просто к увеличению видимых угловых размеров Солнца. Однако максимальная освещенность не может быть больше πB.

Так как диаметр лунной орбиты пренебрежимо мал по сравнению с расстоянием до Солнца, то средняя освещенность поверхности Луны будет такой же, как и у Земли, т.е. $\overline{E}_{\text{С}} = B\sin^2\vartheta_{\text{С}}$. Если Луна как источник света подчиняется закону Ламберта, то ее поверхностная яркость будет $B\sin^2\vartheta_{\text{С}}$. Лучи от Луны, падая на поверхность Земли перпендикулярно, создают освещенность $B\sin^2\vartheta_{\text{С}}\sin^2\vartheta_{\text{Л}}$, где $\vartheta_{\text{Л}}$ — угловой радиус Луны. Эта освещенность много меньше, чем $B\sin^2\vartheta_{\text{С}}$. Поэтому, каковы бы ни были размеры линзы, с помощью лунного света воспламенить папиросу нельзя.

§ 24. Оптические инструменты

1. Лупа. Лупа, как и микроскоп, дает увеличенные изображения предметов, находящихся на небольших расстояниях от глаза. *Увеличение N лупы (и микроскопа) определяется как отношение угла, под которым виден (малый) предмет через лупу, к углу, под которым он был бы виден невооруженным глазом, если бы был помещен от глаза на расстоянии ясного зрения L* (см. § 23, п. 6). Можно также сказать, что увеличение лупы (и микроскопа) есть отношение линейных размеров изображения предмета на сетчатке при рассматривании его в лупу к линейным размерам изображения того же предмета на сетчатке, когда он рассматривается невооруженным глазом с расстояния ясного зрения. При увеличениях, не превосходящих 5, в качестве лупы обычно применяется простая собирательная линза, при больших увеличениях — система, состоящая из таких линз.

При вычислении увеличения лупы будем для общности рассматривать ее как сложную центрированную систему и воспользуемся уравнениями такой системы в форме Ньютона (11.16) и (11.17). Предмет помещается между передним главным фокусом и передней главной плоскостью лупы. Лупа дает прямое мнимое изображение его. Пусть X, Y — координаты точки предмета относительно переднего главного фокуса, X', Y' — координаты соответствующей точки изображения относительно заднего главного фокуса, a — координата глаза (точнее — его передней узловой точки) относительно того же заднего главного фокуса. Расстояние глаза от изображения будет $a - X'$. Глаз через лупу видит отрезок Y' под углом $\varphi' = Y'/(a - X')$. Соответствующий отрезок предмета Y был бы виден невооруженным глазом с расстояния ясного зрения под углом $\varphi = Y/L$. Отсюда

$$N = \frac{\varphi'}{\varphi} = \frac{Y'}{Y}\frac{L}{a - X'},$$

или на основании (11.17)

$$N = \frac{f}{X}\frac{L}{a - X'} = \frac{X'}{f'}\frac{L}{a - X'}.$$

Для лупы $n = n'$, а потому $f' = -f$. Следовательно, с учетом (11.16)

$$N = \frac{L}{f + aX/f}. \tag{24.1}$$

Если глаз помещен в заднем главном фокусе лупы ($a = 0$), то $N = L/f$. То же самое увеличение получится, если предмет поместить в переднем главном фокусе лупы. В этом случае глаз должен быть аккомодирован на бесконечность, т.е. для нормального глаза аккомодирующая мышца должна находиться в расслабленном состоянии. Однако многие наблюдатели ведут наблюдение с напряжением аккомодирующей мышцы, чтобы изображение получалось на привычном для них расстоянии ясного зрения. В этом случае $a - X' = L$, так что

$$N = \frac{X'}{f'} = \frac{a - L}{f'} = \frac{L - a}{f}.$$

Если лупой является тонкая линза, вплотную придвинутая к глазу, то $a = -f$, так что $N = (L/f) + 1$. Рассматриваемый предмет всегда помещается вблизи переднего главного фокуса, т.е. $a/f \ll 1$. Величина aX/f также меньше единицы. Поэтому, независимо от способа наблюдения, увеличение лупы практически всегда определяется формулой

$$N = L/f. \tag{24.2}$$

Подставляя сюда $N = \varphi'/\varphi = (\varphi'/Y)L$, получим $\varphi' = Y/f$. Размер изображения предмета на сетчатке глаза равен $f_{\text{гл}}\varphi' = (f_{\text{гл}}/f)Y$, где $f_{\text{гл}}$ — переднее фокусное расстояние глаза (см. § 21, п. 5). Этот размер определяется практически только фокусным расстоянием лупы f, так как от глаза к глазу величина $f_{\text{гл}}$ меняется незначительно. Таким образом, действие лупы эквивалентно уменьшению расстояния ясного зрения с величины L до f; лупа как бы приближает рассматриваемый предмет к глазу на фокусное расстояние f.

Теоретически увеличение лупы ничем не ограничено. Так, лупа с фокусным расстоянием $f = 0{,}25$ мм имела бы увеличение $N = 1000$. Изготовление луп со столь малыми значениями f, а следовательно и с малыми диаметрами линз, очень затруднительно, а пользование ими крайне неудобно. Кроме того, для таких луп очень велики геометрические и хроматические аберрации. Поэтому увеличение луп, применяемых в действительности, в настоящее время не превосходит 40. Бо́льшие увеличения получаются с помощью микроскопа.

Изображение в лупе формируется с помощью сравнительно тонких пучков лучей, выделяемых зрачком глаза. Поэтому требования к лупам в отношении исправления аберраций не очень жесткие. Хроматическая аберрация положения при малых увеличениях не исправляется вовсе. При более значительных увеличениях и необходимости иметь большое поле зрения надо исправить кому и астигматизм, а также хроматическую разность увеличений. Астигматизм исправляется для главных лучей, проходящих через центр вращения глаза.

Простая плоско-выпуклая линза дает хорошие изображения вплоть до восьмикратного увеличения (т.е. когда ее фокусное расстояние f не менее 3 см). Линза должна быть обращена к глазу плоской стороной. В этом случае, так как предмет помещен вблизи фокуса лупы, преломление лучей происходит практически только на выпуклой поверхности лупы, а не распределяется между обеими поверхностями ее. Это несколько увеличивает сферическую аберрацию. Зато в противоположном случае, когда к глазу обращена выпуклая сторона лупы, значительно увеличиваются аберрации изображений периферийных частей предмета.

2. Микроскоп. Микроскоп состоит из двух собирающих систем линз — *объектива* и *окуляра*, расположенных на большом расстоянии друг от друга. Они помещаются на концах трубки (*тубуса*), которая обычно укрепляется вертикально на штативе микроскопа, объектив ввинчивается в нее снизу, а окуляр вставляется сверху. Предмет помещается на малом расстоянии перед передним фокусом объектива. Объектив дает его действительное обратное увеличенное изображение, которое рассматривается через окуляр, действующий как лупа. Пусть f_1 и f_1' — главные фокусные расстояния объектива, f_2 и f_2' — окуляра. Тогда фокусные расстояния f и f' всей оптической системы микроскопа определятся формулами (12.3), т.е.

$$f = -\frac{f_1 f_2}{\Delta}, \quad f' = \frac{f_1' f_2'}{\Delta}, \tag{24.3}$$

где Δ — оптический интервал между объективом и окуляром. В применении к микроскопу величину Δ называют *оптической длиной тубуса*. Ее следует отличать от *механической длины тубуса* T. Последняя обычно составляет 160 или 190 мм; величина Δ зависит от объектива и окуляра и, как правило, больше T.

Для безиммерсионного микроскопа его увеличение, согласно формуле (24.2), будет

$$N = \frac{L}{f} = -\frac{L}{f_1}\frac{\Delta}{f_2} \tag{24.4}$$

(при наличии иммерсии с показателем преломления n это выражение надо умножить на n). Величину N можно представить в виде $N = N_1 N_2$, где $N_1 = L/f_1$ — увеличение объектива как лупы, $N_2 = \Delta/f_2$ — добавочное увеличение, обусловленное окуляром. При заданном общем увеличении N для устранения ошибок в изображении осевых точек выгоднее делать бóльшим N_1, а для уменьшения ошибок всего поля зрения выгоднее применять бóльшие N_2. По этой причине, а также по соображениям, указанным в п. 1, требуемое общее увеличение N не получают с помощью одних только сильных объективов или одних только сильных окуляров, а идут на компромисс, комбинируя и те и другие.

В микроскопе величина Δ, а также фокусные расстояния f_1 и f_2 положительны. Фокусное расстояние f всей системы, а с ним и увеличение N отрицательны. Это значит, что изображение, получаемое в микроскопе, обратное. Говоря в дальнейшем об увеличении микроскопа, мы обычно будем иметь в виду *абсолютное значение* N.

Величина Δ обычно лежит в пределах 150–200 мм. Фокусное расстояние f_1 объектива при самых больших увеличениях его (100–120) не меньше 1,5 мм, а окуляра — не меньше 8–10 мм. При достаточно больших значениях Δ можно получить увеличение до 2500–3000 раз без чрезмерного уменьшения f_1 и f_2. (Впрочем, столь большие увеличения бесполезны, как это будет подробно разобрано в § 56.)

Штатив микроскопа обычно имеет револьверное приспособление для смены объективов. Такое же приспособление может применяться для смены окуляров. Это позволяет варьировать увеличение микроскопа в широких пределах.

Со стороны предмета в микроскопе велики углы раскрытия, а со стороны изображения — углы наклона лучей. Вследствие этого у объектива особенно велики сферическая аберрация и кома, а у окуляра — астигматизм. Поэтому к объективу и окуляру предъявляются различные требования.

3. Основные требования, предъявляемые к объективу микроскопа: 1) высокая числовая апертура; 2) исправление сферической аберрации для всего очень широкого пучка лучей; 3) выполнение условия синусов для того же пучка; 4) ахроматизация. Ахроматизация состоит в выполнении условия синусов по меньшей мере для двух цветов, а также в получении через объектив ахроматических действительных изображений предмета.

Всем этим требованиям в той или иной мере удается удовлетворить применением довольно сложных конструкций объективов, особенно при больших увеличениях. Наиболее совершенными объективами до настоящего времени являются *апохроматы*, разработанные и рассчитанные Аббе. Впервые апохромат Аббе был выпущен фирмой Цейсса в 1886 г. Это — иммерсионный объектив, состоящий из 10 линз. Он изображен на рис. 90 в увеличенном виде (примерно вдвое). Роль нижней (полусферической) линзы была выяснена в § 18 (п. 4). Поверхность этой линзы погружена в масло, в котором помещен и предмет. Эта линза обладает весьма сильной хроматической аберрацией, уничтожаемой остальными линзами. Из этих остальных линз рассеивающие изготовлены из различных сортов флинта, собирающие — из крона. Аббе нашел, что, перемещая слегка некоторые из этих линз относительно других, можно достигнуть почти полного уничтожения хроматической аберрации. Апохромат ахроматизован для трех и удовлетворяет условию синусов для двух цветов спектра. Фокусное расстояние апохромата Аббе $f = 2$ мм, числовая апертура $n \sin\theta = 1{,}4$, а геометро-оптические качества таковы, что можно считать фактически достигнутыми границы разрешающей способности микроскопа, определяемые волновой природой света (см. § 56). Апохроматы применяются во всех наиболее ответственных микроскопических и микрофотографических исследованиях.

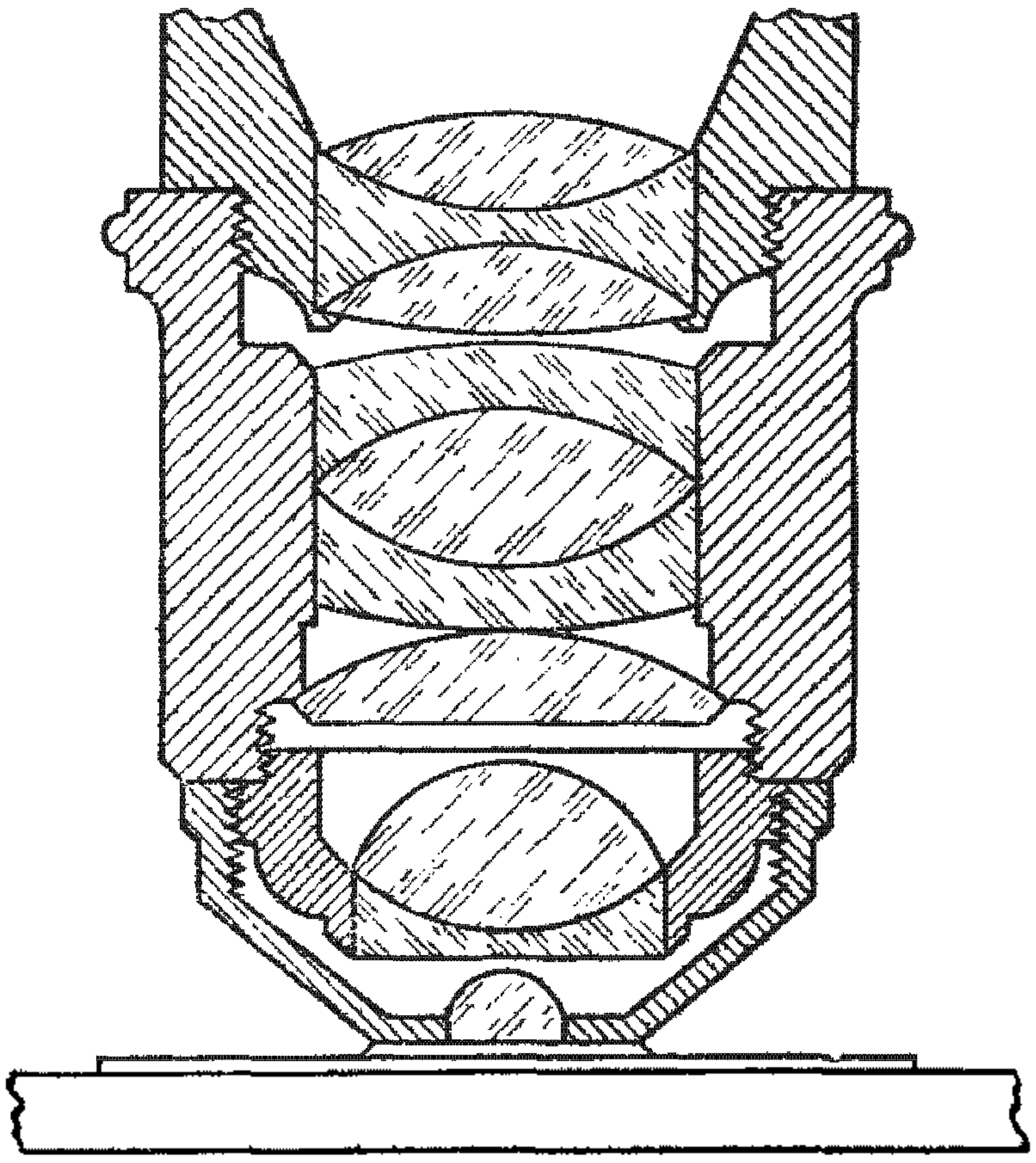


Рис. 90

Менее совершенными, но более распространенными являются *ахроматы*, т.е. объективы, ахроматизованные для двух цветов спектра. Если удовлетвориться частичной ахроматизацией объектива, то надо совместить положения передних фокальных точек для двух цветов (*ахроматизация положения*). Иначе положение изображения предмета, помещенного вблизи передней фокальной точки объектива, очень сильно менялось бы с изменением цвета. Фокусные расстояния при этом остаются неахроматизованными, т.е. неравными. Поэтому хроматическая разность увеличений остается. Она исправляется применением особых окуляров, называемых *компенсационными*. Такие окуляры всегда применяются и в апохроматах, так как у них разность увеличений для лучей различного цвета исправлена не вполне.

Вообще, ахроматизация положения, т.е. совмещение фокусов, или точек схождения лучей различного цвета, должна производиться во всех оптических системах, предназначенных для получения действительных изображений объектов с малыми угловыми размерами. Например, объективы зрительных труб должны быть ахроматизованы в смысле совмещения задних фокальных точек для лучей разного цвета.

Свет от объекта попадает в объектив, проходя через *покровное стекло.* В сухих объективах толщина покровного стекла влияет на сферическую аберрацию. Поэтому все расчеты объективов производят в предположении, что толщина покровного стекла равна 0,17 мм (0,15–0,20 мм). Все сильные сухие объективы в настоящее время снабжены *коррекционной оправой*, с помощью которой можно немного менять расстояние между верхней и нижней линзами объектива и тем самым уничтожать сферическую аберрацию, вносимую покровным стеклом несоответствующей толщины. В объективах с однородной иммерсией, когда показатели преломления покровного стекла, иммерсионной жидкости и фронтальной линзы объектива одинаковы, толщина покровного стекла не имеет значения, так как ее можно компенсировать слоем иммерсионной жидкости между покровным стеклом и объективом.

4. Основные требования, предъявляемые к окуляру: 1) устранение астигматизма в косых пучках; 2) устранение дисторсии; 3) ахроматизация.

Кривизна поверхности изображения обычно невелика и существенно не мешает визуальным наблюдениям. Если объектив дает изображение, свободное от хроматической аберрации, то достаточна уже частичная ахроматизация окуляра. Однако теперь существенна не ахроматизация положения, а *ахроматизация увеличения*, т.е. равенство фокусных расстояний для лучей различного цвета. Действительно, благодаря значительной длине тубуса микроскопа лучи попадают на окуляр под малыми углами наклона к главной оптической оси микроскопа. В этом случае угловое увеличение, даваемое окуляром, определяется формулой (24.2), независимо от того, как аккомодирован глаз: на бесконечность или на расстояние ясного зрения. Если фокусные расстояния окуляра ахроматизованы, то угловое увеличение будет одинаковым для всех цветов. Поэтому изображения предмета на сетчатке глаза во всех цветах совместятся между собой, т.е. окончательное изображение получится неокрашенным, даже если положения главных фокусов и главных плоскостей окуляра не ахроматизованы. Если же изображение, даваемое объективом, не свободно от хроматической аберрации, то ее можно компенсировать хроматической аберрацией противоположного знака, которую дает окуляр. Так, апохроматы дают хроматическую разность увеличений: синее изображение увеличивается сильнее красного. Аббе комбинирует их с такими компенсированными окулярами, у которых фокусное расстояние для синих лучей больше, а следовательно, увеличение меньше, чем для красных лучей.

В § 16 (п. 5) было показано, что фокусное расстояние системы из двух тонких линз, изготовленных из стекла одного и того же сорта, будет одним и тем же для всех цветов спектра, если расстояние l между линзами равно полусумме ее фокусных расстояний: $l = {}^1\!/_2(f_1 + f_2)$. Составлением окуляров из двух линз одновременно можно достигнуть устранения астигматизма косых пучков, а потому окуляры и конструируются по этому принципу. Линза, обращенная к объективу, называется *коллективом*, или *собирательной линзой*, а обращенная к глазу — *глазной линзой*. Коллектив выполняет ту же роль, что и конденсор проекционного аппарата; он собирает все лучи, участвующие в образовании изображения, направляя их в зрачок глаза наблюдателя. При отсутствии специальной диафрагмы поля зрения роль этой диафрагмы выполняет оправа коллектива. Вот почему коллектив часто называют также *полевой линзой*. Увеличение окуляра определяется главным образом глазной линзой. Простейшими окулярами описанного типа, применяемыми в микроскопах и телескопах, являются *окуляр Гюйгенса* и *окуляр Рамсдена* (1735–1800).

5. Окуляр Гюйгенса состоит из двух плоско-выпуклых или двояковыпуклых линз, обращенных более выпуклыми поверхностями в сторону падающих лучей (рис. 91). Для уменьшения сферической аберрации Гюйгенс выбрал фокусные расстояния и взаимное расположение линз так, чтобы отклонение луча, первоначально направленного параллельно главной оптической оси, при преломлении распределялось поровну между обеими линзами. Нетрудно подсчитать (что рекомендуется сделать читателю), что в параксиальном приближении это будет достигнуто при условии $l = f_1 - f_2$, где f_1 — фокусное расстояние собирательной, а f_2 — глазной линзы. Комбинируя это уравнение с условием $l = {}^1\!/_2(f_1 + f_2)$, находим: $f_1 = 3f_2$, $l = 2f_2$. Таким образом, в окуляре самого Гюйгенса $f_2 : l : f_1 = 1 : 2 : 3$ (в современных конструкциях окуляра Гюйгенса чаще пользуются отношением 2 : 3 : 4). Оптический интервал между линзами окуляра $\Delta = -2f_2$. По формулам (12.2), (12.3) и (12.4) нетрудно рассчитать

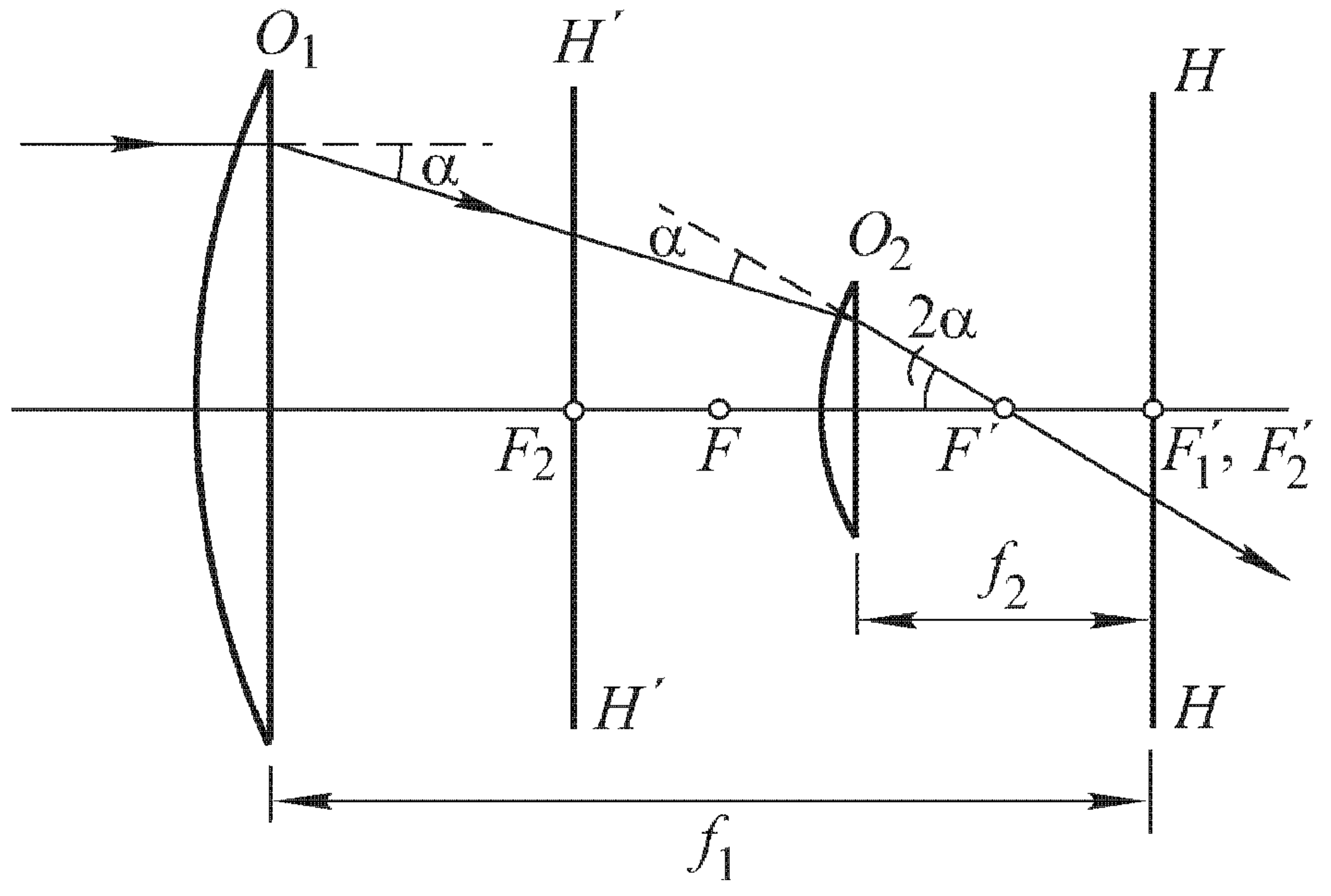


Рис. 91

положение главных фокусов, главных плоскостей и фокусные расстояния окуляра Гюйгенса. Результаты расчета приведены на рис. 91 (положение переднего фокуса F_1 коллектива не показано).

На рис. 92 приведен ход лучей в окуляре Гюйгенса. Предположим, что глаз аккомодирован на бесконечность, а потому окончательное изображение, даваемое микроскопом, должно получаться в бесконечности. Значит, объектив должен давать перевернутое изображение AB предмета в передней фокальной плоскости окуляра. Этого можно добиться, перемещая весь микроскоп вверх или вниз под предметным столиком. Изображение AB для коллектива O_1 играет роль *мнимого предмета*. Коллектив даст его действительное уменьшенное изображение $A'B'$ в передней фокальной плоскости глазной линзы O_2, так как окончательное изображение должно получиться в бесконечности (в этом можно убедиться и непосредственным расчетом). Изображение $A'B'$ увеличивается глазной линзой O_2, действующей как лупа. Угол $A'CB'$ определяет поле зрения. Его можно изменить, поместив в фокальной плоскости F_2 глазной линзы диафрагму поля зрения.

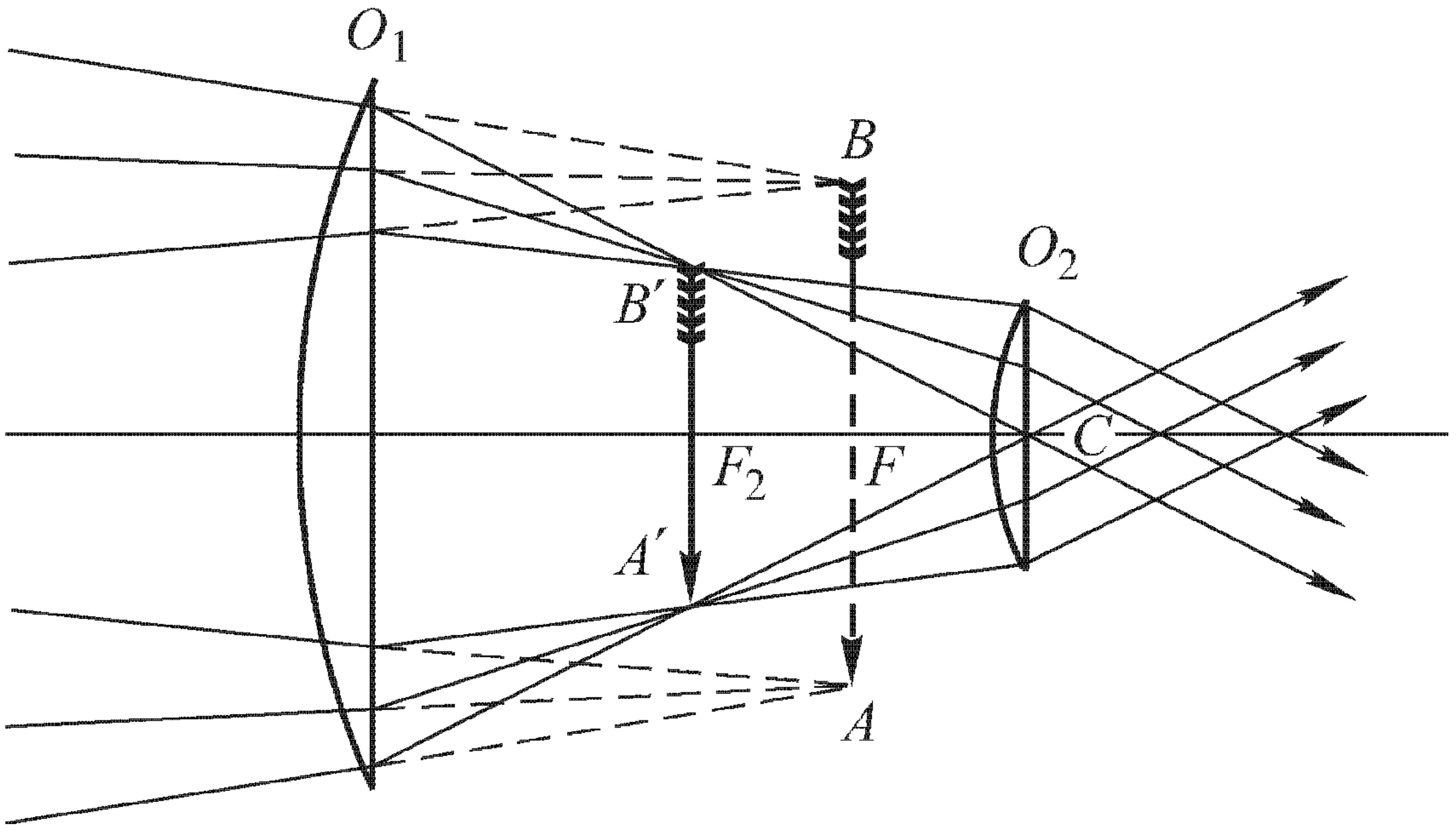


Рис. 92

Окуляр Гюйгенса мало пригоден для точных измерений. Действительно, в этом окуляре шкалу микрометра следовало бы располагать между линзами окуляра. Но тогда шкала микрометра сравнивалась бы с изображением предмета, даваемым объективом микроскопа, которое искажено первой линзой окуляра. Аберрации этой линзы, и прежде всего ее дисторсия, вносили бы в измерения значительные ошибки, не поддающиеся точному учету. Окуляр Гюйгенса не может быть использован и в качестве обыкновенной лупы, предназначенной для непосредственного рассматривания мелких предметов, так как рассматриваемый предмет надо было бы опять помещать между линзами окуляра. Окуляр Гюйгенса применяется в качестве компенсационного окуляра в комбинации с объективами-ахроматами. В этом случае, однако, глазную линзу изготовляют из двух склеенных сферических стекол различного сорта.

6. Окуляр Рамсдена состоит из двух плоско-выпуклых линз с одинаковыми фокусными расстояниями ($f_1 = f_2$), обращенных выпуклыми поверхностями одна к другой (рис. 93). В первоначальном варианте окуляра расстояние между линзами $l = f_1 = f_2$, что обеспечивает ахроматизацию фокусного расстояния f всего окуляра. Но в таком варианте собирательная линза находится в передней фокальной плоскости глазной линзы. Поэтому глазная линза отчетливо изображает каждую пылинку, грязь и царапину на поверхности собирательной линзы, что, конечно, портит изображение предмета. Чтобы это устранить, расстояние между линзами уменьшают, выбирая обычно $l = {}^2\!/_3 f_1$. Это, конечно, нарушает ахроматизацию окуляра. Однако возникающие хроматические явления мало заметны. Их можно устранить, применяя ахроматическую глазную линзу, склеенную из собирательной (крон) и рассеивающей (флинт) линз. Это делается в *окулярах типа Кельнера.* Рассеивающая линза должна быть обращена к глазу. Желательно также ахроматизовать и коллектив окуляра, но этого почти никогда не делают.

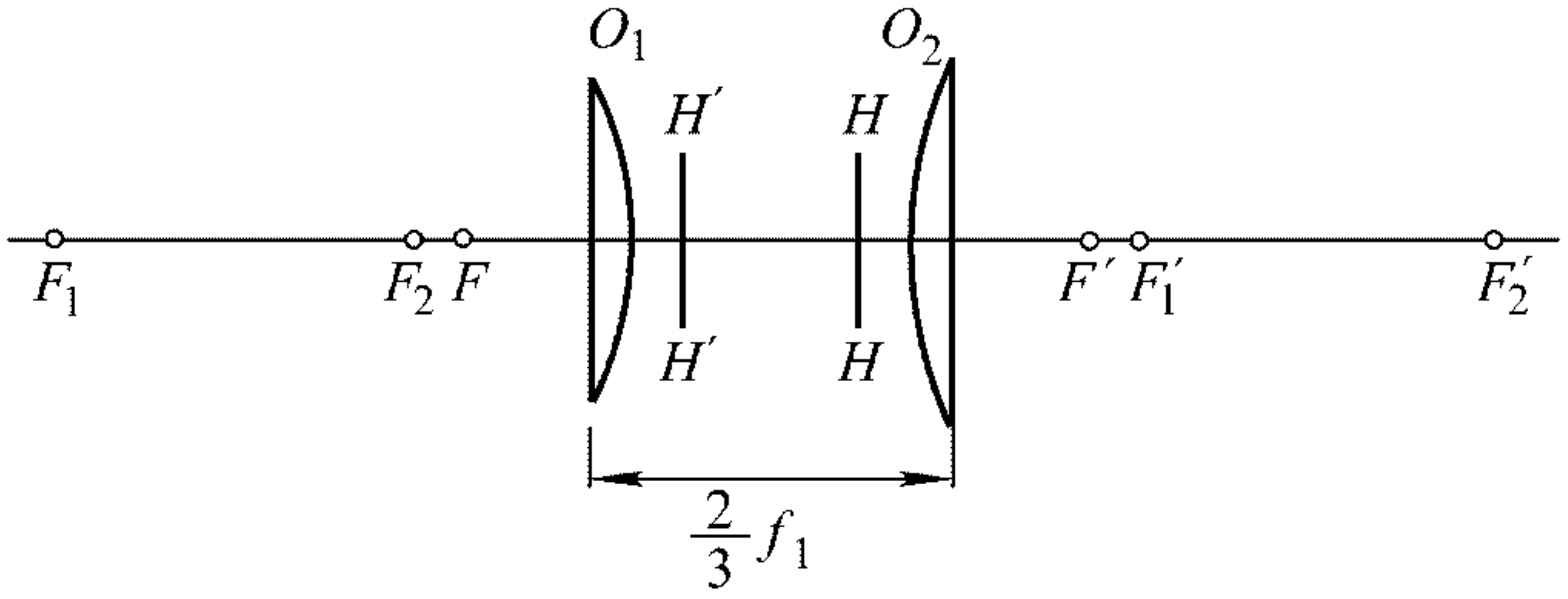


Рис. 93

Положение фокальных точек и главных плоскостей в окуляре Рамсдена показано на рис. 93, а на рис. 94 приведен ход лучей в этом окуляре. Сравнивая их с рисунками 91 и 92, видим невыгоду окуляра Рамсдена: он удлиняет всю систему микроскопа, так как передняя фокальная плоскость окуляра, в которой объектив должен давать действительное изображение AB, лежит перед передней линзой окуляра на расстоянии ${}^1\!/_4 f_1$ (а не между линзами окуляра, как в случае окуляра Гюйгенса).

Но в этой особенности окуляра Рамсдена есть и его преимущество. В фокальной плоскости F можно поместить микрометр (шкалу на стеклянной пластинке или нити, перемещаемые микрометрическим винтом) и таким путем измерить размеры действительного изображения, даваемого объективом микроскопа. При этом микрометр и действительное изображение, даваемое объективом, оба находятся перед окуляром, и аберрации полевой линзы окуляра искажают их одинаково. Тем самым в окуляре Рамсдена устраняется основной недостаток окуляра Гюйгенса как измерительного прибора. Зная линейное увеличение объектива микроскопа, легко вычислить и размеры самого предмета. Поэтому в микроскопах, служащих для измерения длин и углов, применяются окуляры Рамсдена с окулярными микрометрами.

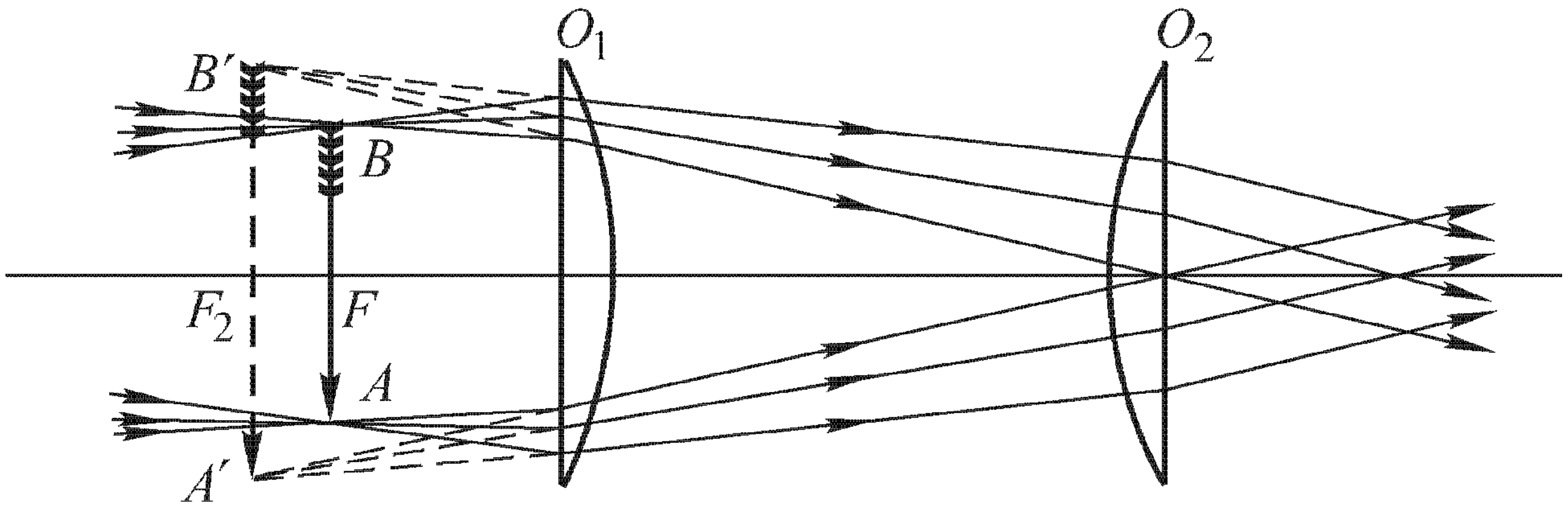


Рис. 94

В зависимости от назначения применяются и более сложные окуляры, состоящие не из двух, а из большего количества линз.

7. На рис. 95 изображен ход лучей в микроскопе с окуляром Гюйгенса. Главные лучи, исходящие из крайних точек предмета AB, начерчены толстыми, а крайние лучи, исходящие из тех же точек, — тонкими линиями. Если увеличение меньше нормального, то зрачок глаза наблюдателя служит апертурной диафрагмой, а его изображение, даваемое всем микроскопом, — входным зрачком этого прибора. Если же увеличение больше нормального, то апертурной диафрагмой и входным зрачком является диафрагма или края полушаровой линзы CD. Ее изображение $C'D'$ будет выходным зрачком микроскопа. Диафрагма поля зрения должна быть помещена в том месте, где получается изображение $A'B'$, даваемое объективом и собирательной линзой окуляра. Изображение $A''B''$, которое видит наблюдатель, можно получить на любом расстоянии, что достигается небольшим поднятием или опусканием всего микроскопа. Обычно его получают на расстоянии ясного зрения (25 см для нормального глаза) или в бесконечности. Перемещая окуляр, можно менять оптическую длину тубуса, а с ней и увеличение микроскопа.

Из рисунка видно главное преимущество микроскопа перед лупой. Оно состоит в том, что в микроскопе задача получения изображения распадается на две части, из которых одна решается объективом, другая окуляром. Объектив дает изображение предмета при максимально возможной числовой апертуре. Окуляр, как и лупа, при большом поле зрения дает изображение точек предмета с помощью узких пучков лучей, так как эти пучки ограничены выходным зрачком микроскопа.

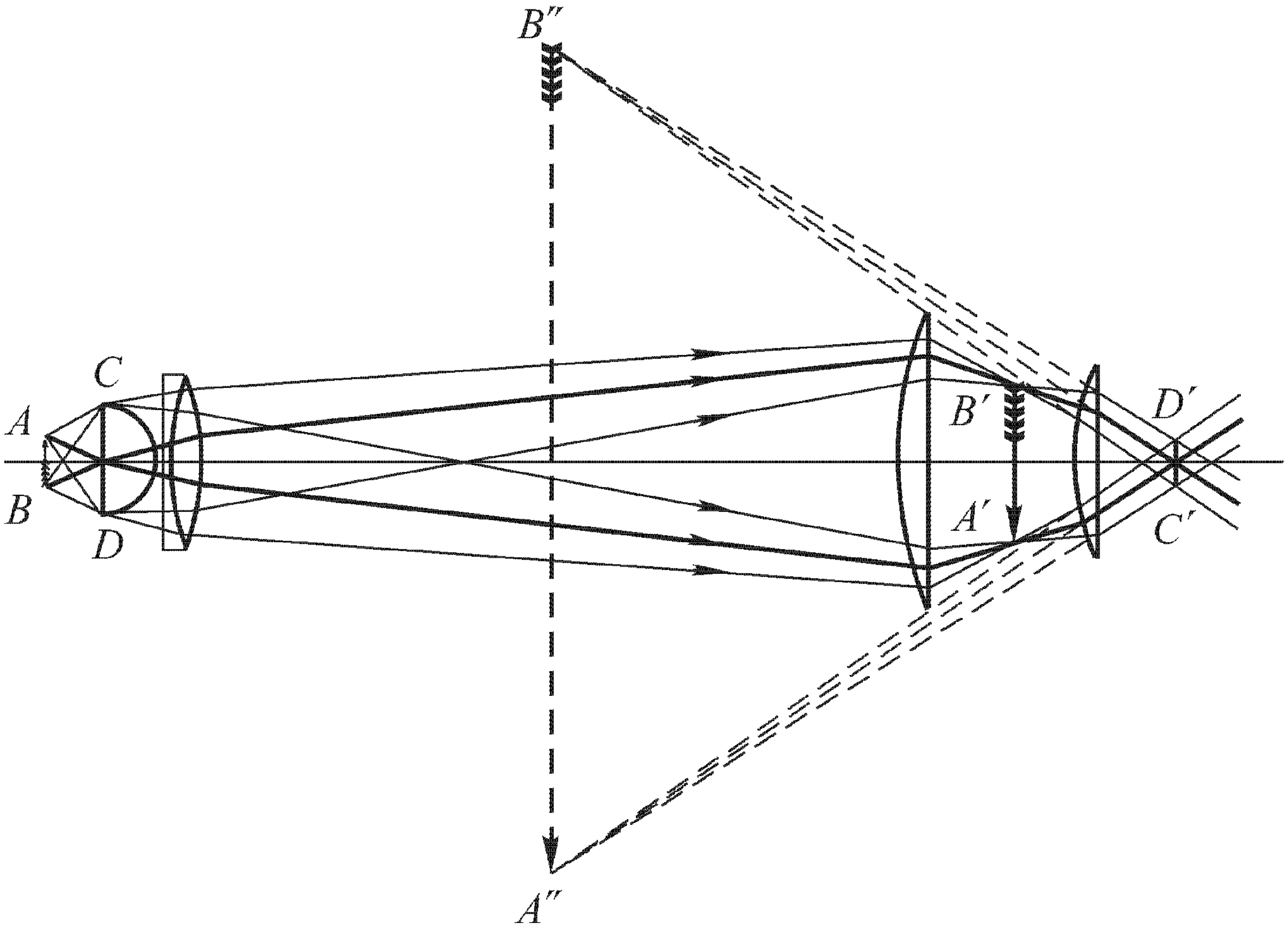


Рис. 95

Объекты, рассматриваемые в микроскоп, обычно не самосветящиеся, а освещаются посторонним источником. Для этой цели употребляется *конденсор*, состоящий из одной или нескольких короткофокусных линз, помещаемый под предметным столиком микроскопа. Чтобы полностью использовать высокую числовую апертуру объектива микроскопа, наклон освещающих лучей к оптической оси должен быть большим. Вот почему для конденсора берут короткофокусные линзы. Конечно, конденсор не может повысить поверхностную яркость источника света, его действие эквивалентно приближению источника к освещаемому объекту.

8. Зрительные трубы и телескопы. Главными частями зрительной трубы являются *объектив* и *окуляр*. Объектив дает в задней фокальной плоскости обратное уменьшенное изображение AB удаленного предмета, рассматриваемое в окуляр как в лупу (рис. 96). Для нормального глаза, аккомодированного на бесконечность, задний фокус объектива должен быть совмещен с передним фокусом окуляра. Это совмещение нарушается, но незначительно, для близорукого и дальнозоркого глаза. Таким образом, параллельный пучок лучей после прохождения через трубу, установленную на бесконечность, остается параллельным, т.е. зрительная труба является *телескопической оптической системой.* Увеличение (угловое) таких систем равно отношению ширины падающего пучка лучей к ширине соответствующего выходящего пучка, или отношению фокусного расстояния объектива f_1 к фокусному расстоянию окуляра f_2 (см.§ 11, п. 10). О нормальном увеличении подробно говорилось в п. 4 предыдущего параграфа.

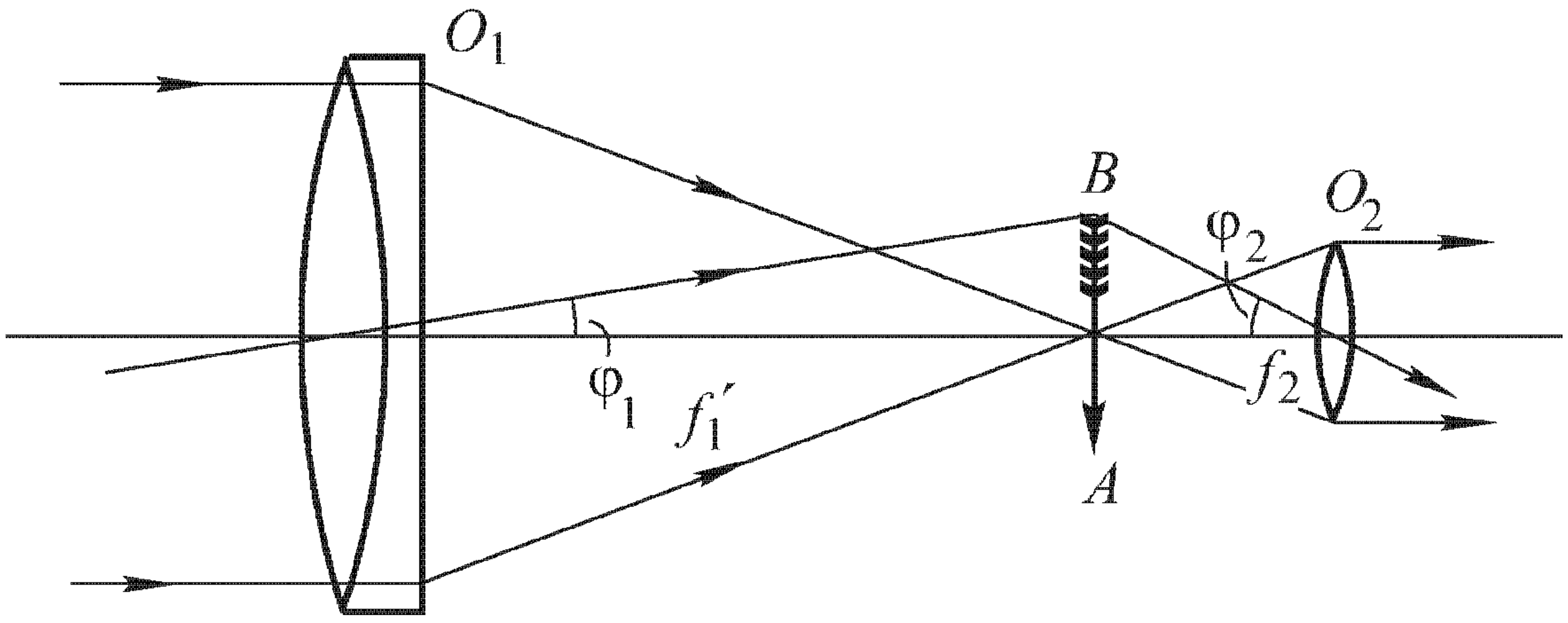


Рис. 96

Лучи, падающие на объектив зрительной трубы, наклонены под малыми углами к главной оптической оси. Поэтому нет надобности в тщательном исправлении аберраций, зависящих от этих углов. Объективом служит ахроматическая двойная линза, исправленная на сферическую аберрацию. Кома, т.е. отступление от условия синусов, может быть исправлена не столь тщательно. Ахроматизация окуляра должна удовлетворять тем же условиям, что и в случае микроскопа: если главный луч падает на окуляр под небольшим углом к главной оптической оси, то можно ограничиться ахроматизацией только фокусного расстояния. Поэтому для микроскопов и зрительных труб можно пользоваться одними и теми же окулярами. Если требуются микрометрические измерения изображения (как в астрономических трубах), то предпочтительнее пользоваться окуляром Рамсдена, а не Гюйгенса.

На рис. 96 изображен ход лучей в зрительной трубе Кеплера (для простоты сложный окуляр заменен простой собирательной линзой). Эта труба дает перевернутое изображение. Чтобы устранить этот недостаток, за изображением AB, даваемым объективом, на двойном фокусном расстоянии помещают добавочную ахроматизованную собирательную линзу, которая переворачивает изображение AB, не меняя его величины. Полученное изображение и рассматривается в окуляр. Окончательное изображение, которое видит глаз, получается прямым.

В трубе Галилея окончательное изображение получается прямым без введения промежуточной линзы. Это достигается тем, что в качестве окуляра применяется простая двояковогнутая линза, передний фокус которой F_2 совпадает с задним фокусом объектива F_1' (рис. 97). Поле зрения галилеевой трубы невелико. Зато при том же объективе галилеева труба значительно короче кеплеровой. Поэтому труба с отрицательным окуляром применяется только в театральных биноклях.

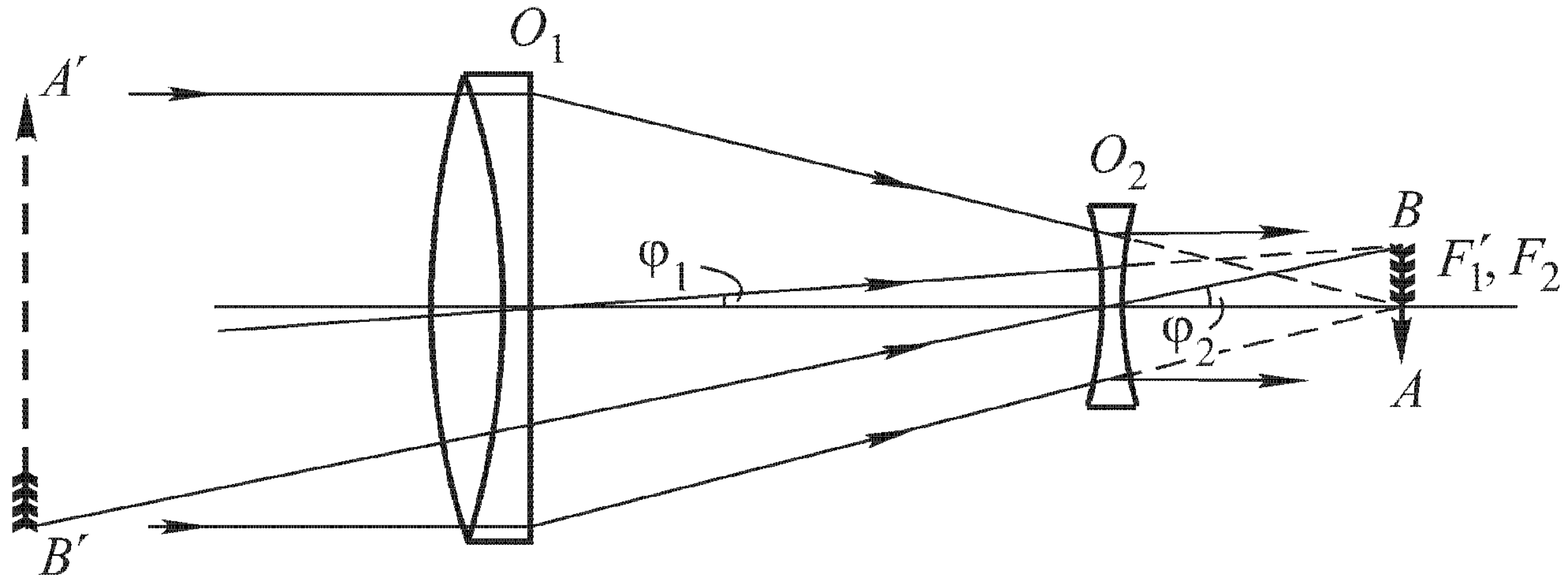


Рис. 97

В обыкновенном бинокле, состоящем из двух зрительных труб, переворачивание изображения, даваемого объективом, достигается в каждой трубе с помощью четырехкратного полного отражения от гипотенузных поверхностей двух прямоугольных стеклянных призм A и B, расположенных, как показано на рис. 98. Луч выходит из призмы B параллельно падающему лучу, но получает боковое смещение. В остальном устройство зрительной трубы не отличается от устройства трубы Кеплера. Так как луч проходит расстояние между призмами трижды, то можно значительно уменьшить длину трубы, раздвигая призмы A и B друг относительно друга. При несколько иной конструкции можно получить бинокулярную трубу, у которой расстояние между объективами во много раз превышает расстояние между глазами наблюдателя. В такой «стереотрубе» сильно увеличивается *стереоскопический эффект*, обусловленный зрением двумя глазами.

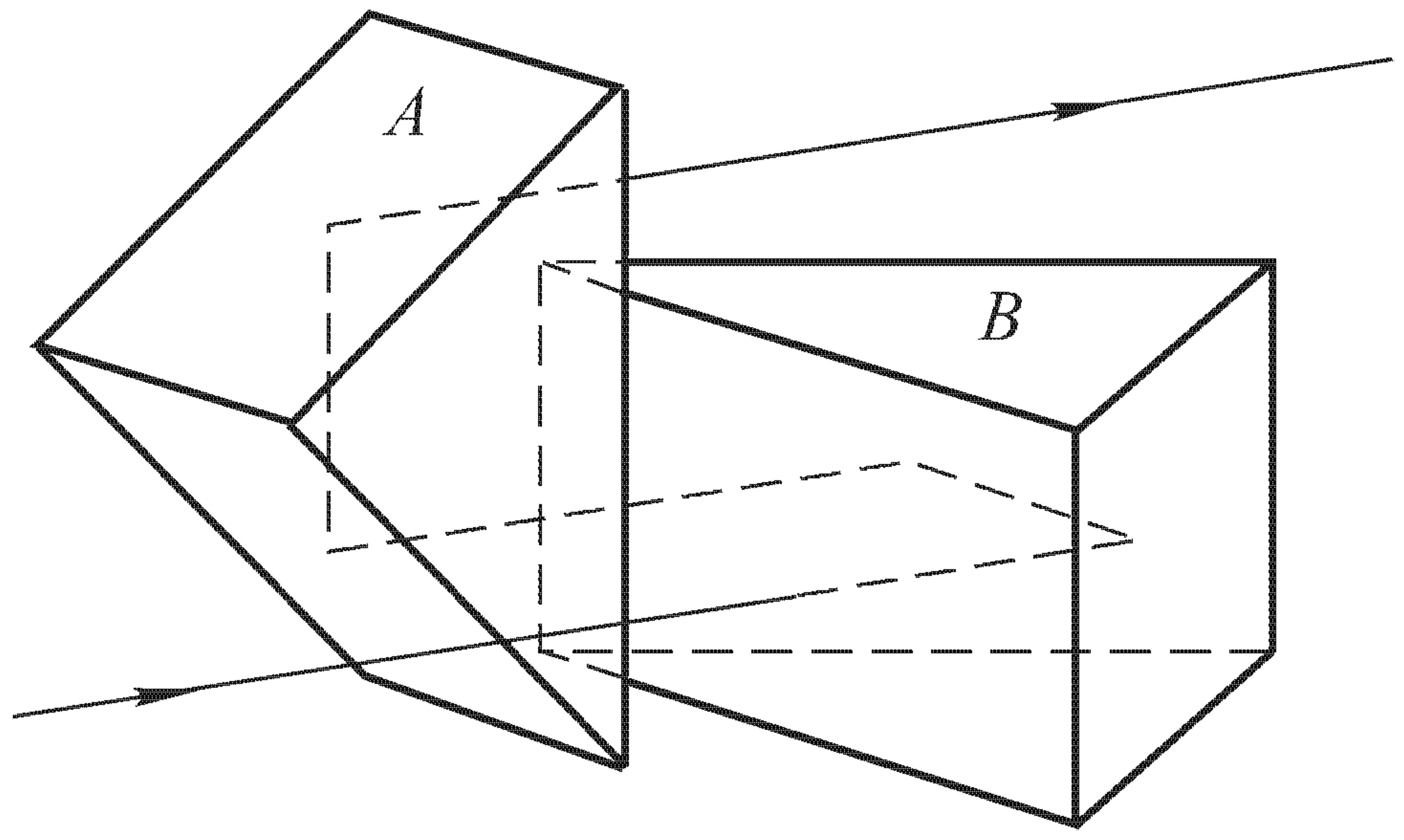


Рис. 98

9. *Астрономические трубы*, или *телескопы*, бывают двух типов: преломляющие (*рефракторы*) и отражающие (*рефлекторы*). В преломляющих объективом служит линза, в отражающих — вогнутое зеркало, называемое *главным зеркалом*.

Изготовление объективов для телескопов-рефракторов представляет большие трудности, быстро возрастающие с увеличением размеров объектива. Главная трудность состоит в получении большого диска стекла с вполне однородным показателем преломления и отсутствием натяжений. Последние снимаются медленным *отжигом* стекла, длящимся месяцами. Необходима большая точность в изготовлении сферических поверхностей, в центровке и сборке линз. Влияние неизбежных небольших случайных отклонений от полной однородности показателя преломления стекла устраняется деформацией точных сферических поверхностей объектива путем *местной полировки* — *ретуши*. Изготовление объектива для рефракторов больших обсерваторий тянется годами. Самым крупным рефрактором в настоящее время является телескоп Иеркской обсерватории в США. Диаметр его объектива составляет 40 дюймов (1 м). Диаметр объектива рефрактора Главной астрономической обсерватории АН СССР в Пулкове равен 30 дюймов (75 см).

Значительно легче изготовить главное зеркало телескопа, так как свет не проходит сквозь стекло, и последнее может быть менее высокого качества, но должно быть лишь хорошо отожжено. Такое стекло может быть отлито в диски огромного размера. Поэтому крупнейшими являются отражательные телескопы (рефлекторы). Самый крупный телескоп с диаметром главного зеркала 6 м недавно установлен в СССР на северных отрогах Кавказского хребта вблизи станицы Зеленчукской. До этого самым крупным был телескоп в обсерватории Маунт Паломар (Калифорния, США) с диаметром главного зеркала 5 м.

Отражающая поверхность главного зеркала телескопа-рефлектора шлифуется на стекле. Ей придают форму параболоида, чтобы лучи, параллельные главной оптической оси, собирались в фокусе параболоида. При изготовлении поверхности зеркала сначала придают форму сферы, радиус которой R почти в точности равен удвоенному фокусному расстоянию F параболоида. Затем производят параболизацию этой поверхности путем снятия шлифовкой (ретушью) тонкого слоя стекла на центральной части зеркала. В наиболее толстом месте, как нетрудно подсчитать, толщина слоя должна составлять ${}^1\!/_8(r/R)^4R$, где r — радиус кривизны зеркала. При $r = 50$ см, $F = 5$ м для этой толщины получаем $7{,}8 \cdot 10^{-4}$ см, т.е. примерно 15 длин волн желто-зеленого света. Отражающая поверхность зеркала покрывается тонким слоем серебра, алюминия или родия. С течением времени под действием воздуха отражательная способность металлического слоя постепенно уменьшается. Серебряный слой должен обновляться примерно каждые полгода. Алюминиевый слой более устойчив и держится годами. Еще лучшие результаты дает покрытие зеркала слоем родия.

Важное преимущество зеркального телескопа перед рефрактором состоит в том, что этот телескоп абсолютно ахроматичен. Это позволяет не только работать со светом любой окраски, но и значительно повысить светосилу телескопа с соответствующим уменьшением его длины.

Свет должен дважды пройти трубу рефлектора, прежде чем он соберется в фокусе. Поэтому требуется небольшое вспомогательное зеркало (или призма), чтобы сместить фокальную плоскость в более удобное положение для наблюдения глазом или помещения фотографической пластинки. В зеркальном телескопе Ньютона плоское вспомогательное зеркало s смещало фокус F вбок, как указано на рис. 99 (на этом и последующих рисунках окуляр не изображен, указано только смещенное положение F фокуса оптической системы).

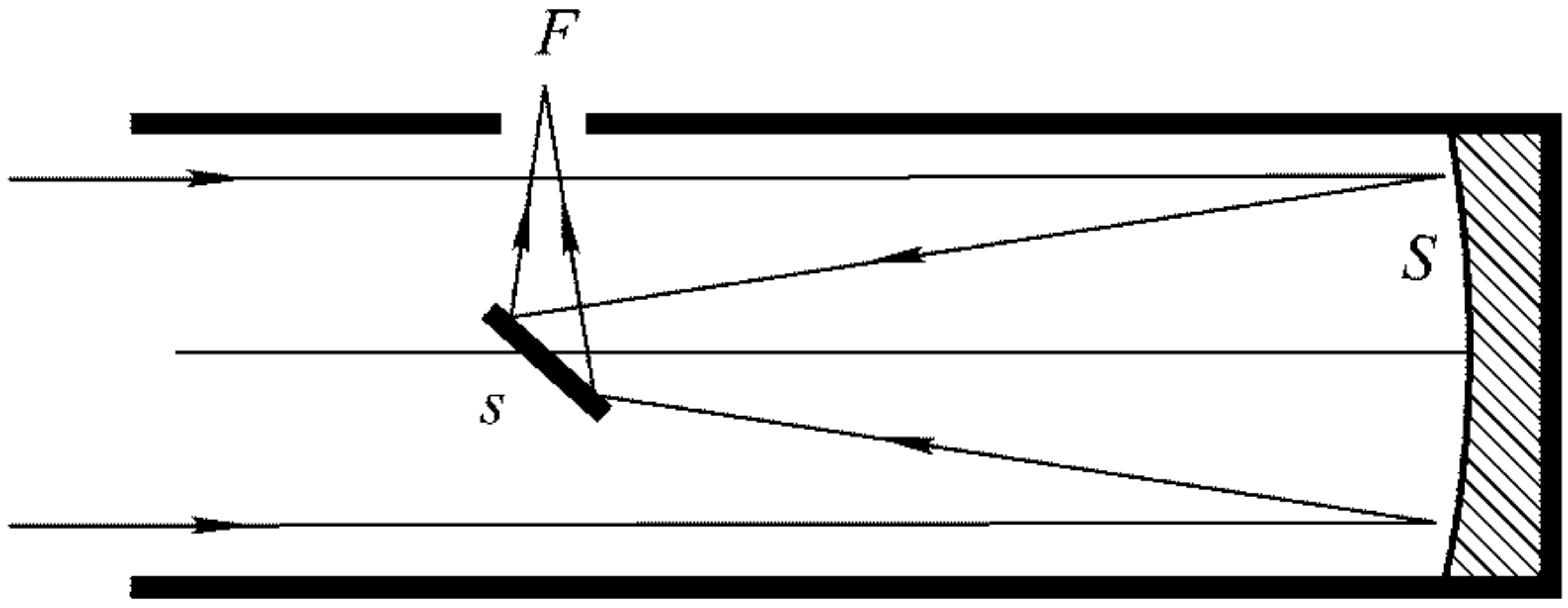


Рис. 99

В телескопе Вильяма Гершеля (рис. 100) вспомогательного зеркала нет, главный фокус F смещается вбок трубы путем небольшого наклона главного зеркала S. За несколько лет до Гершеля такой телескоп был изобретен Ломоносовым, но его изобретение оставалось неизвестным вплоть до начала XX века. Недостаток системы Ломоносова–Гершеля состоит в том,

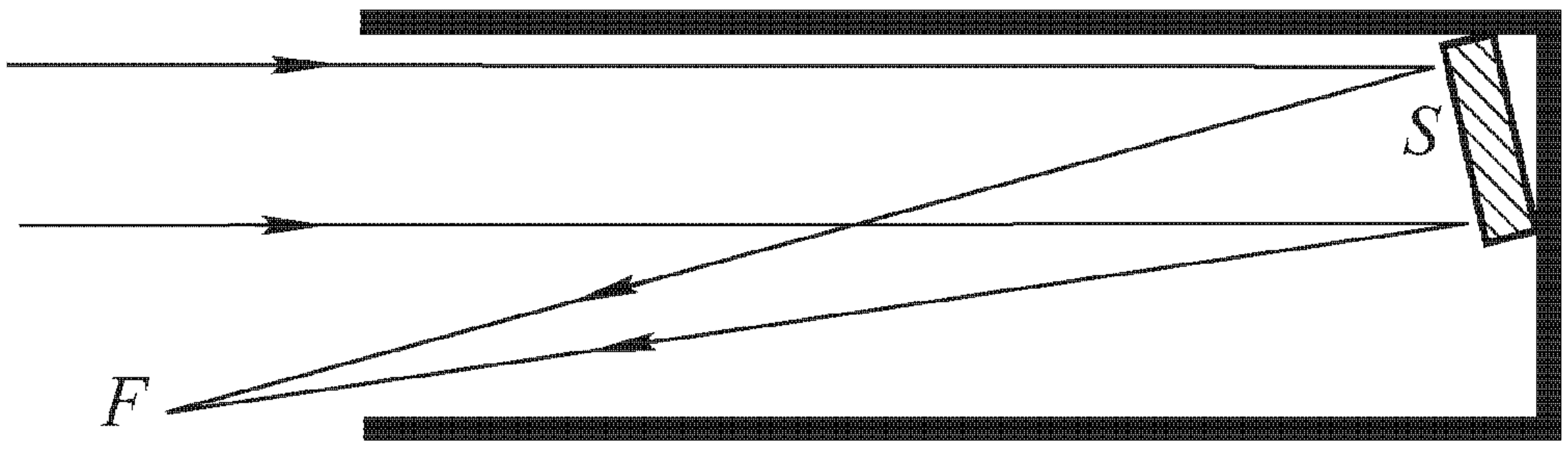


Рис. 100

что изображение получается не на главной оптической оси. Это ухудшает качество изображения. Впрочем, сам Гершель пользовался столь длиннофокусными зеркалами, что требовались лишь небольшие углы наклона, и ухудшение изображения было незначительным.

В системе Кассегрена (рис. 101) применяется выпуклое гиперболическое зеркало s, располагающееся на главной оптической оси несколько ближе фокуса главного зеркала S. Фокус гиперболической поверхности вспомогательного зеркала s совпадает с фокусом главного зеркала S. После отражения лучей от зеркала образуется новый фокус F оптической системы, совпадающий со вторым фокусом гиперболической поверхности зеркала s. Лучи выходят из системы через отверстие в центре главного зеркала S и попадают в окуляр или на фотопластинку.

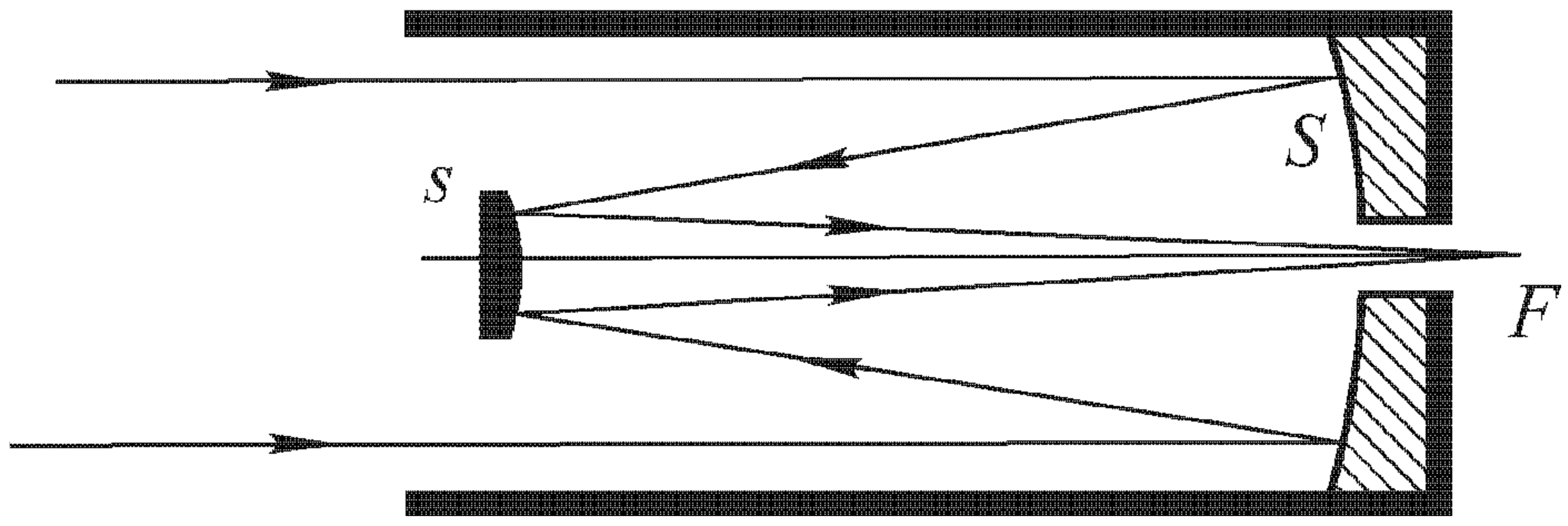


Рис. 101

Сходное устройство имеет телескоп Грегори (рис. 102). Здесь вспомогательным зеркалом s служит вогнутое эллиптическое зеркало, располагающееся на главной оптической оси несколько дальше фокуса главного зеркала S. Фокус зеркала S совмещается с первым фокусом эллиптического зеркала s. Смещенный фокус F всей оптической системы получается в точке нахождения второго фокуса эллиптической поверхности зеркала s. Лучи также выходят из системы через отверстие в центре главного зеркала S и попадают в окуляр или на фотопластинку. По качеству изображения телескоп Грегори несколько превосходит телескоп Кассегрена. Но для него требуется гораздо более длинная труба. По этой причине система Грегори почти никогда не применяется.

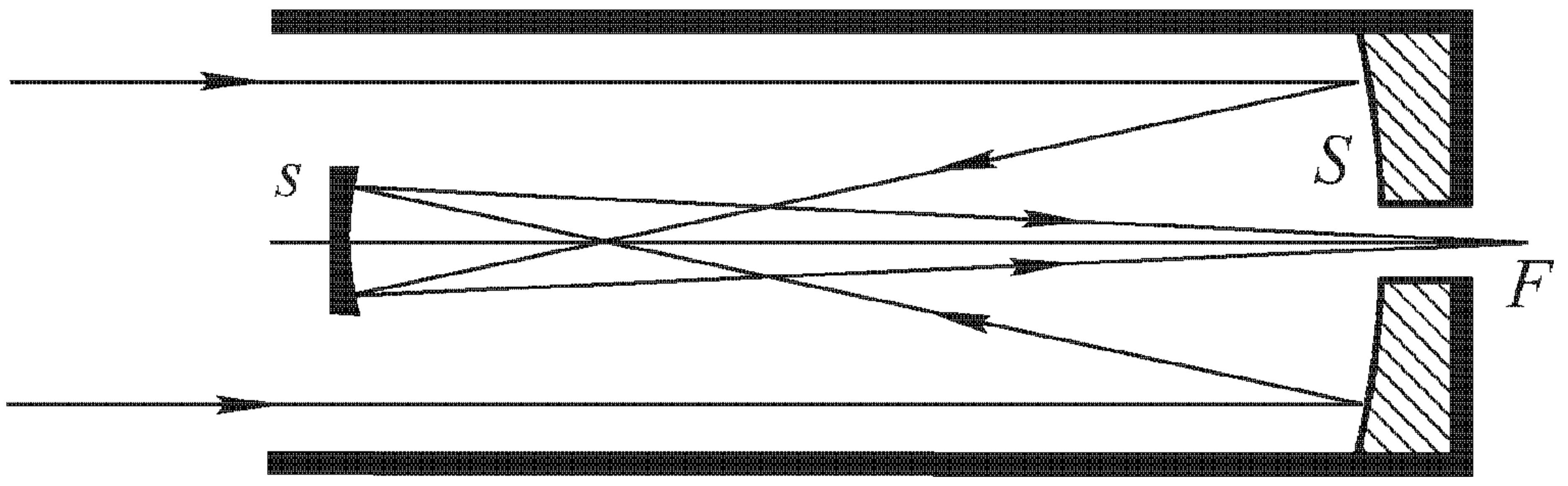


Рис. 102

10. Камера Шмидта и менисковые системы Максутова. Придание отражающей поверхности главного зеркала телескопа параболической формы, а также использование в качестве вспомогательных эллиптических и гиперболических зеркал устраняет сферическую аберрацию, но сохраняет все прочие геометрические аберрации, так как геометрические фокусы параболоида, эллипсоида и гиперболоида являются *только анаберрационными, но не апланатическими точками.* Зеркальным объективам телескопов всегда свойственны значительные аберрации комы и астигматизма. Вследствие этого поле зрения, где получаются четкие изображения, у этих приборов невелико и измеряется минутами, а в лучших случаях — десятками минут. В 1930 г. Б. Шмидт, сотрудник Гамбургской обсерватории в Бергедорфе, предложил новый тип телескопа, получивший название камеры Шмидта. Короткофокусная камера Шмидта с относительным отверстием $D/f = 1$ может давать резкие изображения при поле зрения $\approx 25°$. Параболическое же зеркало при таком же относительном отверстии может иметь полезное поле зрения, измеряемое только несколькими дуговыми минутами.

Рисунок 103 поясняет основную идею камеры Шмидта. На этом рисунке S_1S_2 — вогнутое сферическое зеркало с центром в точке C, а D_1D_2 — отверстие диафрагмы с центром в той же точке, введенной только для лучшего уяснения указанной идеи. Пучок лучей A, параллельных главной оптической оси зеркала, после отражения пройдет через кружок с центром в главном фокусе F. Размеры кружка определяются сферической аберрацией. Параллельный пучок A_1, падающий наклонно, пройдет через аналогичный кружок с центром в побочном фокусе F_1. Геометрическим местом всех таких фокусов будет сфера, радиус которой вдвое меньше радиуса зеркала S_1S_2. Она изображена штриховой линией. Таким образом, бесконечно удаленные точки в результате отражения света от сферического зеркала S_1S_2 изобразятся кружками, расположенными на этой штриховой сфере.

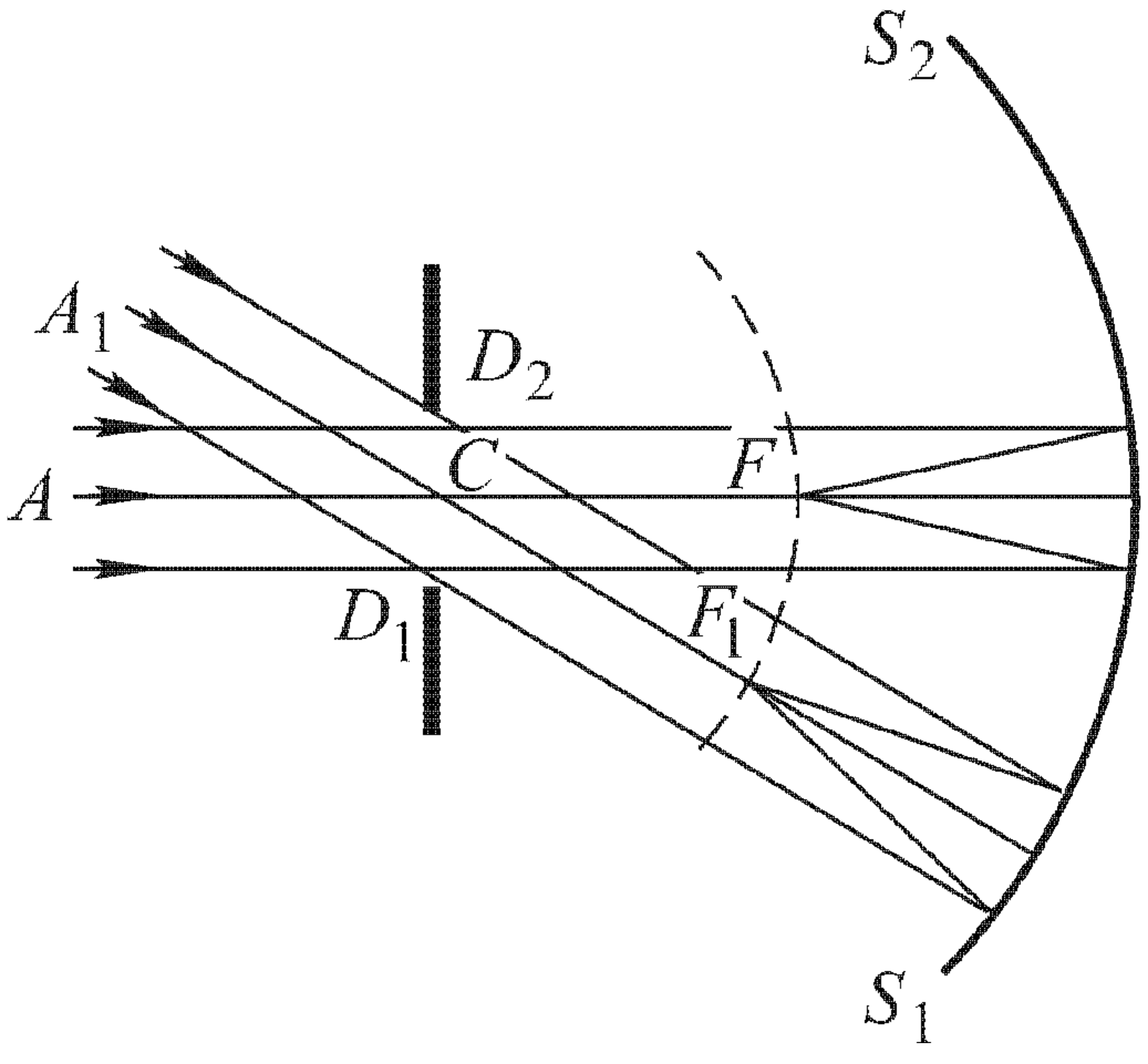


Рис. 103

Размеры кружков почти одинаковы, так как они образуются параллельными пучками лучей практически одинакового поперечного сечения. Наклон пучков влияет лишь на положение кружков на штриховой сфере, но не на их величину. Впрочем, полная одинаковость кружков несущественна. Существенно, что единственными геометрическими аберрациями будут сферическая аберрация и искривление поверхности изображения. Комы, астигматизма и дисторсии не будет.

В обычных зеркальных телескопах сферическая аберрация устраняется параболизацией отражающей поверхности зеркала. Но такой способ годится только для точек, лежащих на главной оптической оси зеркала. Шмидт предложил устранять сферическую аберрацию коррекционной стеклянной пластинкой, помещаемой перед сферическим зеркалом S_1S_2. Одна поверхность пластинки плоская, вторая — поверхность вращения надлежащей формы, мало отклоняющаяся от плоскости. Это отклонение настолько мало, что оно совершенно незаметно невооруженному глазу.

Сечение коррекционной пластинки P_1P_2 представлено на схематическом рис. 104 в преувеличенном виде. Центральная часть пластинки более толстая — действует как собирательная линза, приближая фокус центральных лучей к зеркалу. Края пластинки более тонкие и, подобно рассеивающей линзе, удаляют от зеркала фокусы наклонных пучков лучей.

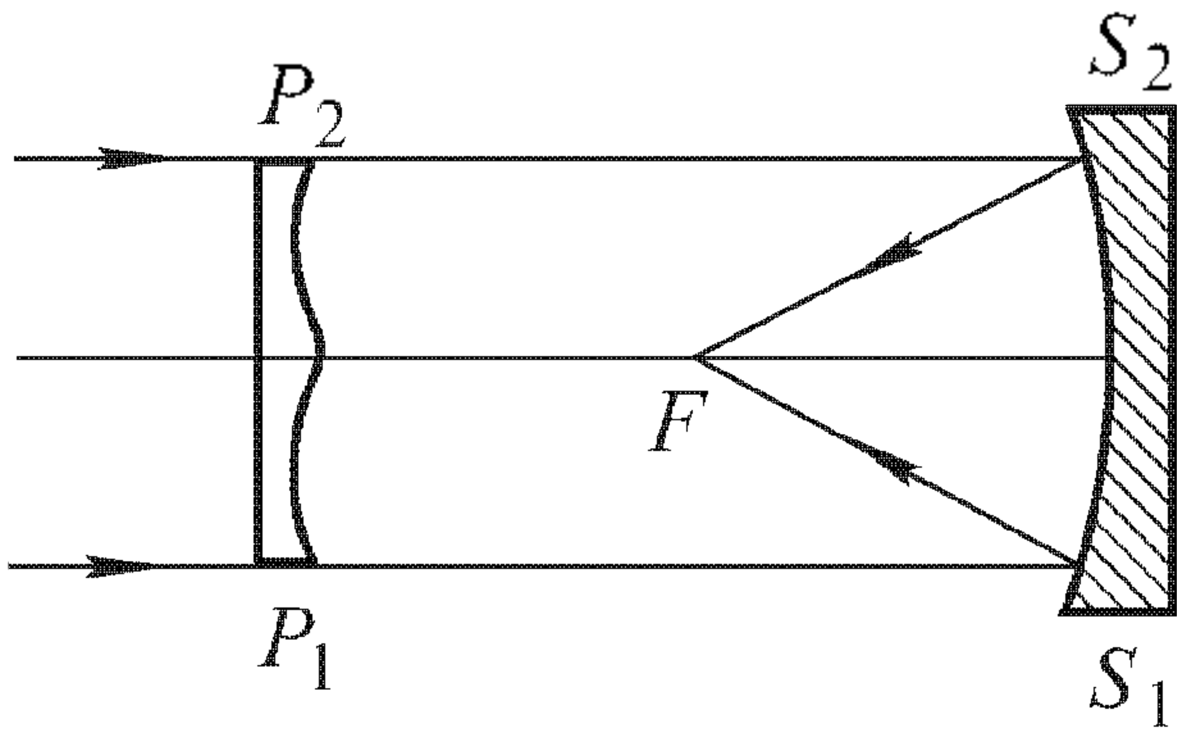


Рис. 104

Помещение коррекционной пластинки вплотную к зеркалу не может дать никаких преимуществ по сравнению с параболизацией его поверхности. Однако коррекционная пластинка, если правильно рассчитать ее форму, помещенная на определенном расстоянии от зеркала, устраняет сферическую аберрацию не только для пучков лучей, параллельных главной оптической оси, но и для пучков, наклоненных к ней. Обычно коррекционную пластинку помещают в центре сферического зеркала и для этого положения рассчитывают ее форму. При этом другие аберрации практически не возникают. Изготовление коррекционных пластинок Шмидта — очень трудная задача.

В 1941 г. Д.Д. Максутов (1896–1964) для исправления аберраций сферического зеркала предложил применять *мениски*, значительно более легкие в изготовлении, чем коррекционные пластинки Шмидта с несферическими поверхностями. Системы Максутова быстро получили широкое распространение. При определенном соотношении между радиусами кривизны сферических поверхностей мениска и его толщиной мениск ахроматизован (для двух требуемых цветов спектра).

На рис. 105 *а* приведена простейшая система Максутова, состоящая из сферического зеркала S и мениска B, обращенного к нему выпуклой стороной. Если мениск ахроматизован, то в фокусе F образуется ахроматическое изображение. Сферическая аберрация мениска подбирается так, чтобы компенсировать сферическую аберрацию зеркала S. При правильно выбранном расстоянии l между мениском и зеркалом оказывается исправленной и кома, т.е. система становится апланатической. Если при том же расстоянии l мениск повернуть к зеркалу вогнутой стороной, то апланатизм нарушается. Его можно восстановить, увеличивая расстояние l (рис. 105 *б*). Первая ориентация мениска более выгодна, так как при ней телескоп получается вдвое короче, чем при второй. Однако для некоторых целей предпочтительнее вторая ориентация.

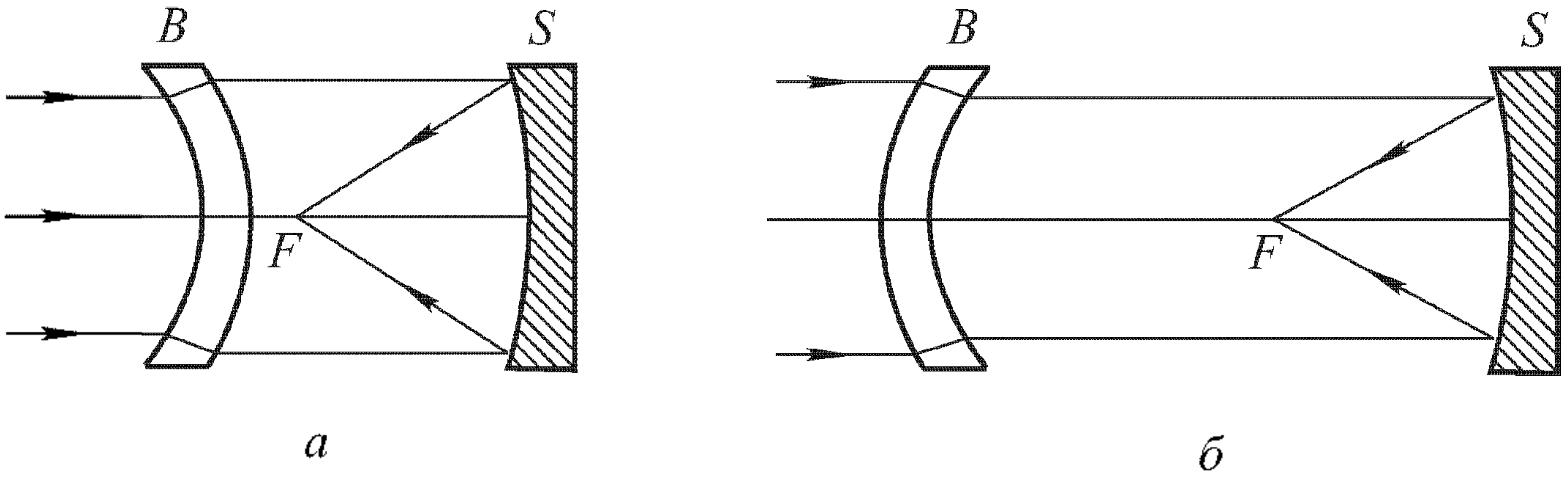


Рис. 105

Астигматизм менисковой системы может быть сделан ничтожно малым, а в частных случаях и вполне устранен. Искривление поверхности изображения остается. Но из всех аберраций эта аберрация является наименее вредной, и существуют способы ее устранения. Соединение камеры Шмидта или менисковых систем Максутова с окуляром производится так же, как в телескопах Ньютона, Кассегрена и Грегори.

§ 25. Электрические и магнитные линзы

1. Допустим, что показатель преломления n меняется в пространстве *непрерывно*. Проведем поверхности равного показателя преломления и притом настолько часто, что показатели преломления между каждыми соседними поверхностями можно будет считать величинами постоянными. Тогда непрерывное изменение величины n заменится скачкообразным, происходящим на границах между слоями. Если среда обладает осевой симметрией, то эти границы будут поверхностями вращения, вершины которых лежат на оси симметрии системы. В малой окрестности вокруг оси симметрии их можно аппроксимировать сферами, центры которых также лежат на той же оси. Таким путем мы приходим к центрированной системе тонких сферических линз, у которой ось симметрии служит главной оптической осью и к которой применимы все результаты оптики параксиальных лучей. Увеличивая число слоев бесконечно и одновременно устремляя к нулю их толщины, мы восстановим в пределе первоначальное непрерывное распределение показателя преломления. Отсюда следует, что *осесимметричную среду с непрерывно изменяющимся в пространстве показателем преломления можно рассматривать как предельный случай центрированной системы линз и применять к ней законы и методы оптики параксиальных лучей. Такая среда обладает способностью давать оптические изображения.*

Оптические системы с непрерывно изменяющимися показателями преломления принципиально возможны, но из-за трудностей их изготовления в световой оптике они не встречаются. (Исключение составляет хрусталик глаза, показатель преломления которого возрастает от периферии к центру.) Аналогом таких систем являются электронные и ионные приборы (электронный микроскоп, электронный осциллограф, электронно-лучевая трубка в телевидении и пр.), в которых роль лучей играют электроны или ионы, движущиеся в электростатических или магнитных полях, создаваемых заряженными электродами или катушками, по которым текут электрические токи. Эти электроды называются *электрическими*, а катушки — *магнитными линзами*. Получение изображений в таких системах изучается в *электронной и ионной оптике.*

2. Возможность фокусировки заряженных частиц и получения изображений в осесимметричных системах электростатических линз непосредственно следует из формальной аналогии между геометрической оптикой и классической механикой, отмеченной нами в п. 1 § 4. Действительно, согласно этой аналогии, траектория частицы в потенциальном силовом поле совпадает со световым лучом в непрерывной среде, показатель преломления которой пропорционален скорости (а в релятивистской механике — импульсу) частицы, однозначно определяющейся ее положением в пространстве. Например, можно положить

$$n = \sqrt{2(W - eV)}, \tag{25.1}$$

где W — полная энергия частицы, e — ее заряд, V — потенциал электростатического поля.

3. Несколько труднее понять действие магнитных линз и комбинаций таких линз с электростатическими линзами. Для исследования вопроса рассмотрим статическое электромагнитное поле, обладающее симметрией вращения вокруг некоторой оси. Эту ось примем за ось X цилиндрической системы координат. Расстояние до оси X будем обозначать через r, а азимутальный угол — через φ. Ввиду симметрии вращения, электрическое и магнитное поля не могут зависеть от φ. Предположим кроме того, что всякая плоскость, проходящая через ось X, является плоскостью симметрии поля. Тогда не будет φ-составляющих электрического и магнитного полей ($E_\varphi = B_\varphi = 0$). Оставшиеся x- и r-составляющие этих полей будут функциями только координат x и r. Уравнение $\operatorname{div}\mathbf{B} = 0$ после умножения на r запишется в виде

$$\frac{\partial}{\partial x}(rB_x) + \frac{\partial}{\partial r}(rB_r) = 0. \tag{25.2}$$

Введем потенциальную функцию $A(x, r)$, определяемую условием $rB_x = -\partial A/\partial r$. Тогда предыдущее уравнение преобразуется в

$$\frac{\partial}{\partial r}(rB_r) = \frac{\partial}{\partial x}\left(\frac{\partial A}{\partial r}\right) = \frac{\partial}{\partial r}\left(\frac{\partial A}{\partial x}\right),$$

откуда $rB_r = \partial A/\partial x$, причем «постоянную интегрирования» (точнее — функцию x) мы без ущерба для общности положили равной нулю. Таким образом,

$$rB_x = -\frac{\partial A}{\partial r}, \quad rB_r = \frac{\partial A}{\partial x}. \tag{25.3}$$

Отсюда

$$A(x,r) = \int r(B_r\,dx - B_x\,dr). \tag{25.4}$$

В силу (25.2) входящий сюда интеграл не зависит от пути интегрирования, а только от его начальной и конечной точек. Условимся помещать начальную точку на оси системы. На оси системы в силу симметрии $B_r = 0$. Поэтому не имеет значения, в каком месте этой оси выбрать начальную точку. От этого значение интеграла (25.4) не зависит, и он может быть представлен в виде

$$A(x,r) = -\int\limits_0^r B_x r\,dr. \tag{25.5}$$

Обратимся теперь к исследованию движения частицы с массой m и зарядом e, ограничиваясь ради простоты нерелятивистским приближением. Угловую скорость $\dot\varphi$ частицы можно найти из уравнения моментов

$$\frac{d}{dt}(mr^2\dot\varphi) = M_x,$$

где M_x — момент сил, действующих на частицу, относительно оси X. Электрическое поле не влияет на момент M_x. Он создается только силой $\mathbf{F} = \frac{e}{c}[\mathbf{vB}]$, действующей со стороны магнитного поля. Вычисляя этот момент, находим

$$M_x = \frac{er}{c}(B_r v_x - B_x v_r) = \frac{e}{c}\left(v_x\frac{\partial A}{\partial x} + v_r\frac{\partial A}{\partial r}\right).$$

Если под x, r и φ понимать координаты движущейся частицы, то $A(x,r)$ станет функцией времени t. Производная этой функции равна

$$\frac{dA}{dt} = \frac{\partial A}{\partial x}\frac{dx}{dt} + \frac{\partial A}{\partial r}\frac{dr}{dt} = v_x\frac{\partial A}{\partial x} + v_r\frac{\partial A}{\partial r}.$$

Следовательно,

$$M_x = \frac{e}{c}\frac{dA}{dt}, \quad \frac{d}{dt}(mr^2\dot\varphi) = \frac{e}{c}\frac{dA}{dt}.$$

Отсюда

$$\dot\varphi = \frac{e}{mcr^2}A + \frac{C}{mr^2}.$$

Постоянная C должна обращаться в нуль. Иначе (так как на оси системы функция A равна нулю) при $r = 0$ мы получили бы $r\dot\varphi = \infty$. А это невозможно, так как $r\dot\varphi$ есть линейная скорость частицы в ее вращении вокруг оси X. Таким образом,

$$\dot\varphi = \frac{eA}{mcr^2}. \tag{25.6}$$

Возьмем теперь уравнение движения частицы

$$m\mathbf{a} = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{vB}]\right), \tag{25.7}$$

где $\mathbf{a}$ — ускорение. При проецировании его на ось X и направление радиуса r слева получим соответственно $ma_x = m\ddot x$ и ma_r. Радиальное ускорение в цилиндрической системе координат определяется выражением $a_r = \ddot{r} - \omega^2 r$ (см. т. I, § 46). Введем еще электрический потенциал V и учтем соотношения (25.3) и (25.6). Тогда после недлинных преобразований из (25.7) получим

$$m\ddot{x} = -e\frac{\partial}{\partial x}\left(V + \frac{eA^2}{2mc^2r^2}\right), \quad m\ddot{r} = -e\frac{\partial}{\partial r}\left(V + \frac{eA^2}{2mc^2r^2}\right). \tag{25.8}$$

Задача о движении частицы распалась на две *независимые задачи*: определение угловой координаты φ и определение координат x и r. Для решения первой задачи имеется уравнение (25.6), а для второй — уравнения (25.8). Формально вторая задача идентична с задачей определения траектории заряженной частицы в плоском электростатическом поле с потенциалом

$$U = V + \frac{eA^2}{2mc^2r^2}. \tag{25.9}$$

Ее формальным аналогом в оптике служит задача о распространении светового луча в неоднородной изотропной среде с показателем преломления

$$n = \sqrt{2m(W - eV) - \frac{e^2A^2}{c^2r^2}} = \sqrt{n_0^2 - \frac{e^2A^2}{c^2r^2}}, \tag{25.10}$$

где $n_0 = \sqrt{2m(W - eV)}$ — «показатель преломления» в отсутствие магнитного поля. Для получения окончательной формы траектории частицы надо наложить на этот «луч» *дополнительное вращение*, выражаемое формулой (25.6).

Учтем теперь условие параксиальности. Для этого разложим $B_x(x, r)$ в ряд по степеням r. Ввиду осевой симметрии, это разложение может содержать только четные степени r. Оборвем разложение на члене нулевой степени — в этом приближении поле B_x не зависит от r. Вынося B_x из-под знака интеграла (25.5) и выполнив интегрирование, получим

$$A = -\frac{1}{2}B_x r^2, \tag{25.11}$$

а после подстановки в формулу (25.6)

$$\dot{\varphi} = -eB_x/(2mc). \tag{25.12}$$

Отсюда

$$\frac{d\varphi}{dx} = \frac{d\varphi}{dt}\frac{dt}{dx} = -\frac{1}{v_x}\frac{eB_x}{2mc}. \tag{25.13}$$

В параксиальном приближении скорость v_x также не зависит от расстояния r до оси X и равна скорости частицы, движущейся вдоль этой оси. Поэтому производные $d\varphi/dx$ имеют одинаковые значения для всех частиц, независимо от наклона их траекторий к оси X. Если l — расстояние от предметной плоскости до плоскости изображения, то на этом расстоянии в параксиальном приближении все частицы поворачивают вокруг оси симметрии системы на один и тот же угол

$$\varphi = -\frac{e}{2mc}\int_0^l \frac{B_x}{v_x}\,dx. \tag{25.14}$$

Следовательно, *в смысле получения изображений система будет вести себя так, как если бы магнитного поля не было, а электростатическое* *определялось потенциалом* (25.9). *Магнитное поле приводит еще к несущественному повороту всего изображения вокруг оси симметрии системы на угол, определяемый формулой* (25.14). В высших приближениях угол поворота зависит от наклона траектории к оси системы. Это ведет к появлению дополнительных аберраций, обусловленных наличием магнитного поля.

4. Отметим еще одно обстоятельство. Если электрического поля нет, а магнитное поле однородно, то частица движется по спирали, вращаясь с циклотронной частотой $\omega = -eB/mc$ (см. т. III, § 86). Между тем формула (25.12) для этого случая дает угловую скорость, вдвое меньшую ω. Недоразумение легко разъясняется, если заметить, что ω есть угловая скорость вращения частицы *вокруг оси спирали*, тогда как формула (25.12) определяет ее вращение *вокруг одной из образующих* той же спирали. Допустим, например, что магнитное поле перпендикулярно к плоскости рисунка (рис. 106), а частица вращается по окружности в той же плоскости. За ось X можно, конечно, принять любую прямую, направленную вдоль поля **B**. Поэтому можно считать, что частица выходит из какой-то точки M этой оси. Описав окружность, она снова возвращается в исходную точку. Ее вращение вокруг точки M описывается формулой (25.12), тогда как вращение вокруг центра O окружности происходит с циклотронной частотой ω. За время dt частица повернется вокруг центра O на угол $\omega\,dt$, а вокруг точки M — на угол $d\varphi = {}^1\!/_2\omega\,dt$, так что $\dot{\varphi} = {}^1\!/_2\omega$.

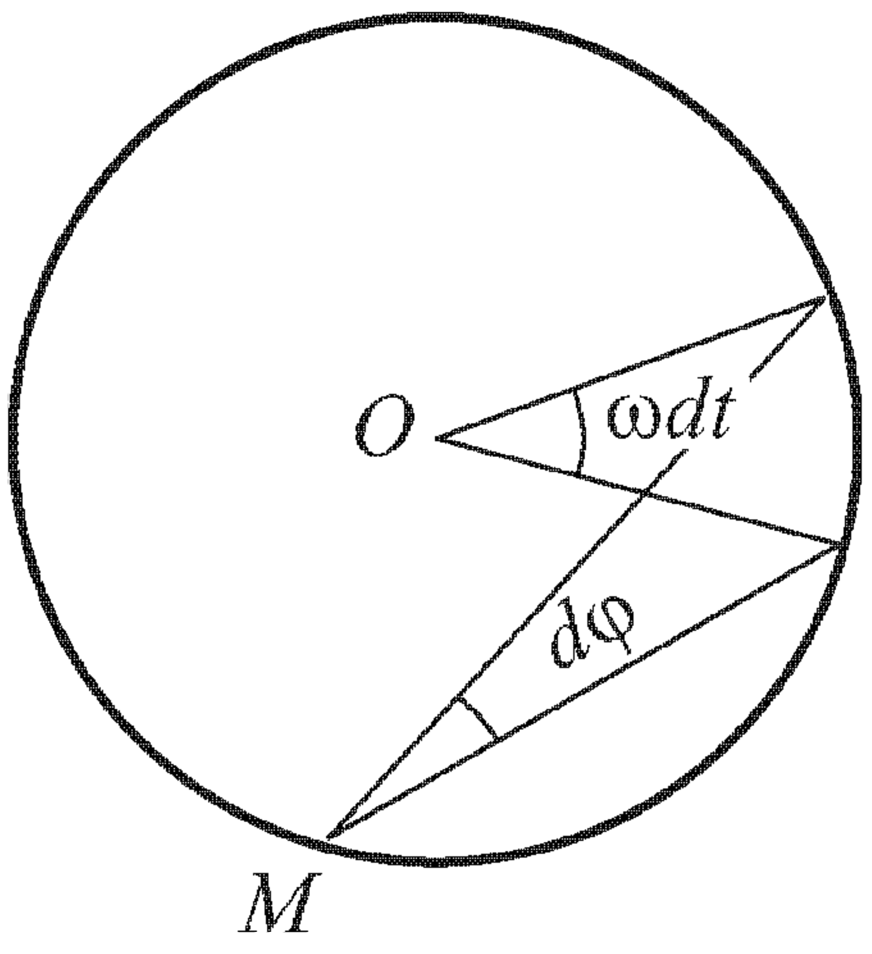


Рис. 106

5. Выше фокусировка частиц в электрических линзах была объяснена с помощью аналогии со световой оптикой. Сделаем теперь то же самое, рассматривая силы, действующие на частицу. На рис. 107 представлена линза, состоящая из трех соосных металлических цилиндров одинакового диаметра. Крайние цилиндры заземлены, на средний подан положительный или отрицательный потенциал. Линзы такого типа используются в электронно-лучевых трубках и некоторых электронных микроскопах. На рисунке изображены электрические силовые линии с указанием направлений сил, действующих на частицу.

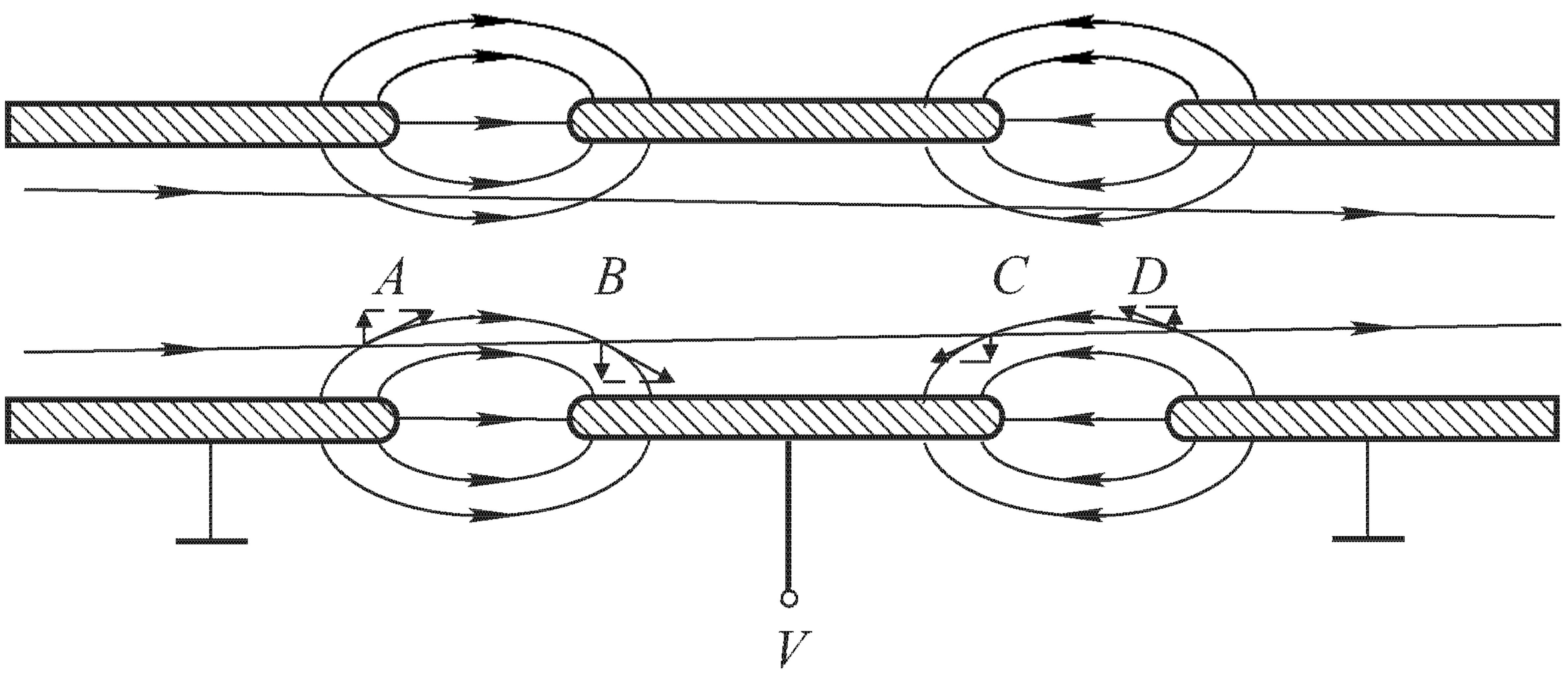


Рис. 107

Допустим, что частица влетает в линзу, двигаясь параллельно ее оси. В области A действующая на нее сила имеет составляющую, направленную вверх. Эта сила будет смещать частицу вверх. В области B направление вертикальной составляющей силы изменится на противоположное. Однако, так как под действием электрического поля скорость частицы непрерывно возрастает, на прохождение области B частица затрачивает меньше времени, чем на прохождение области A. Поэтому поперечная скорость, приобретенная частицей в области A, не может быть скомпенсирована скоростью противоположного направления, которую она получает в области B. В результате в областях A и B и по выходе из них частица будет двигаться вверх, приближаясь к оси линзы. Аналогично, в области C на частицу действует сила, стремящаяся удалить ее от оси линзы, а в области D — приблизить. Но в этих областях частица замедляется, а потому проводит в области D большее время, чем в C. Поэтому при прохождении обеих областей C и D вертикальная скорость частицы, направленная вверх, возрастет.

Эти разъяснения объясняют, почему частицы приближаются к оси линзы. Конечно, из них не следует, что все частицы пучка соберутся в одной и той же точке на оси линзы. Для доказательства этого требуется уже количественное рассмотрение, которое и было проведено выше на основе аналогии со световой оптикой.

ЗАДАЧИ

1. Вычислить фокусное расстояние тонкой оптической линзы, показатель преломления которой в пространстве меняется непрерывно.

Решение. Сначала не будем вводить предположение о тонкости линзы, а рассмотрим среду, обладающую симметрией вращения вокруг оси X. Уравнение луча в меридиональной плоскости представится в виде $r = r(x)$. Обозначим через u угол, образуемый касательной к лучу с осью X (рис. 108). В параксиальном приближении квадратом этого угла пренебрегают. В этом приближении кривизна луча определяется выражением $1/R = -du/dx$, причем радиус кривизны R мы считаем положительным, когда луч обращен вогнутостью к главной оптической оси X, и отрицательным в противоположном случае. Воспользуемся формулой (4.1). В пределах точности параксиальной оптики $\partial r/\partial N = -\cos u \approx -1$, $\partial x/\partial N = \sin u \approx u$, так что

$$\frac{\partial n}{\partial N} = \frac{\partial n}{\partial r}\frac{\partial r}{\partial N} + \frac{\partial n}{\partial x}\frac{\partial x}{\partial N} \approx -\frac{\partial n}{\partial r} + u\frac{\partial n}{\partial x}.$$

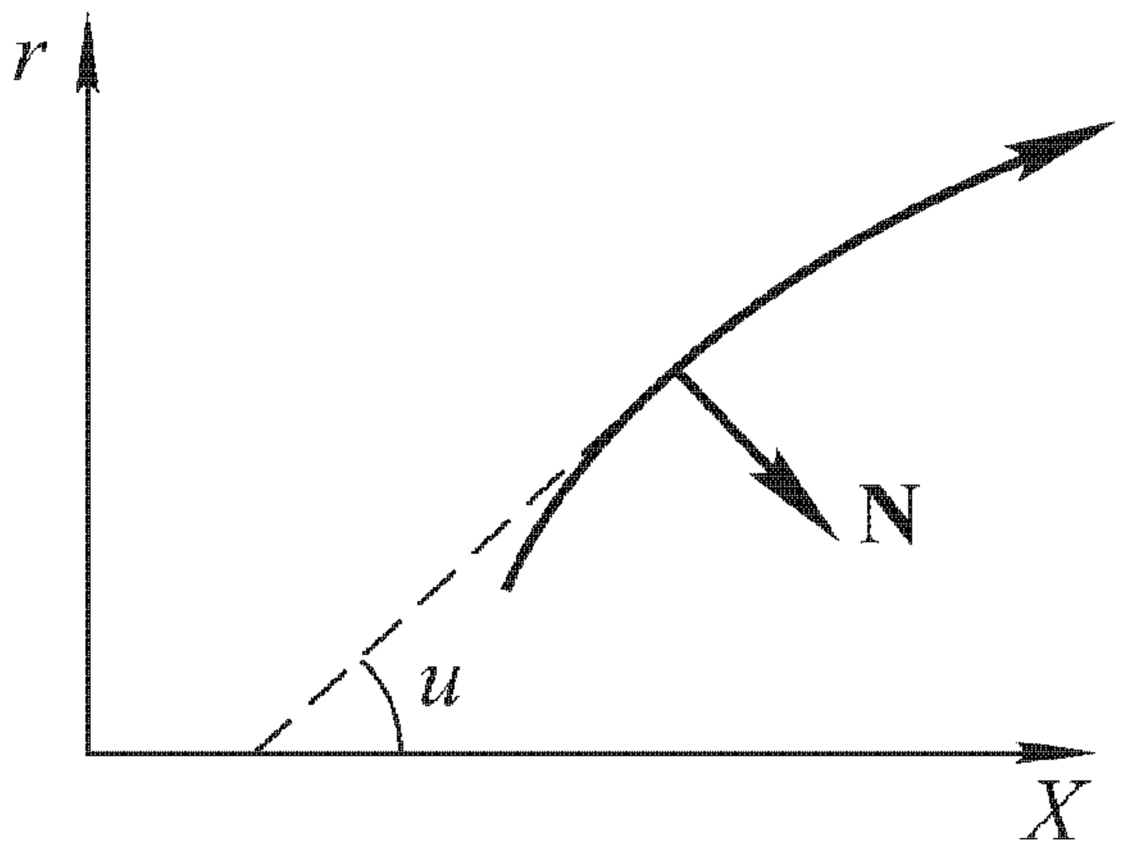


Рис. 108

Так как на луче $r = r(x)$, то $n(x,r)$ можно рассматривать как сложную функцию от x, т.е. $n(x) = n[x, r(x)]$. Ее производная по x определяется выражением

$$\frac{dn}{dx} = \frac{\partial n}{\partial x} + \frac{\partial n}{\partial r}\frac{\partial r}{\partial x} = \frac{\partial n}{\partial x} + u\frac{\partial n}{\partial r},$$

т.е. отличается от частной производной $\partial n/\partial x$ на величину первого порядка малости по u. Поэтому замена $\partial n/\partial x$ на dn/dx в предыдущем выражении вносит ошибку второго порядка, которой мы пренебрегаем. Сделав эту замену и воспользовавшись формулой (4.1), получим

$$-\frac{1}{R}=\frac{du}{dx}=-\frac{1}{n}\frac{\partial n}{\partial N}=\frac{1}{n}\frac{\partial n}{\partial r}=-u\frac{dn}{dx},$$

или

$$\frac{d}{dx}(nu)=\frac{\partial n}{\partial r}. \tag{25.15}$$

Разложим $n=n(x,r)$ в ряд по степеням r и оборвем разложение на члене второй степени. Член первой степени должен отсутствовать, ввиду осевой симметрии системы. Таким образом,

$$n=n_0(x)+\frac{1}{2}\left(\frac{\partial^2 n}{\partial r^2}\right)_0 r^2,$$

а потому

$$\frac{\partial n}{\partial r}=r\left(\frac{\partial^2 n}{\partial r^2}\right)_0.$$

Здесь нуль в индексе означает, что величина с таким индексом берется при $r=0$, т.е. на главной оптической оси системы. В результате уравнение (25.15) переходит в

$$\frac{d}{dx}(nu)=r\left(\frac{\partial^2 n}{\partial r^2}\right)_0. \tag{25.16}$$

Допустим теперь, что линза тонкая. Пусть P и P' — оптически сопряженные точки на ее оси (рис. 109). Отрезки соединяющего их луча вне линзы прямолинейны. Проинтегрируем уравнение (25.16) по x в пределах от $-\infty$ до $+\infty$. Фактически это сводится к интегрированию по отрезку AB, так как вне этого отрезка $\partial^2 n/\partial r^2=0$. Внутри линзы, поскольку она тонкая, радиус r можно считать постоянным и вынести его из-под знака интеграла. Это дает

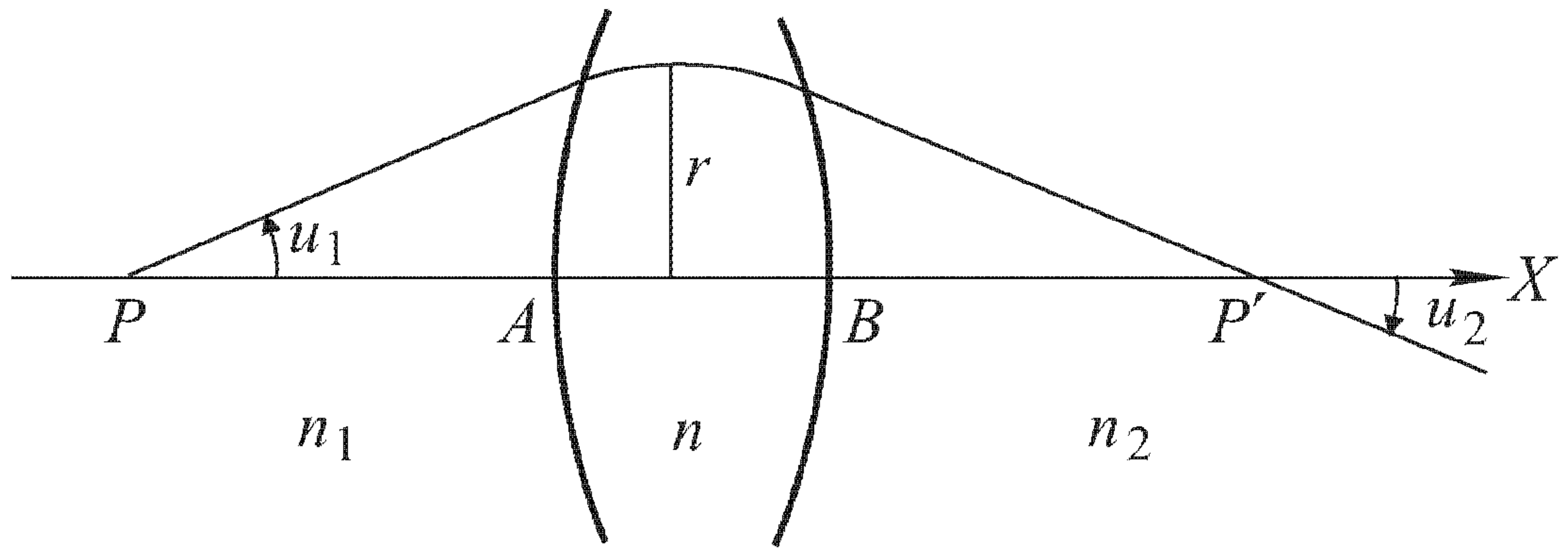


Рис. 109

$$n_2u_2-n_1u_1=r\int\limits_{-\infty}^{+\infty}\left(\frac{\partial^2 n}{\partial r^2}\right)dx,$$

где n_1 — показатель преломления в пространстве предметов, а n_2 — в пространстве изображений. Обозначим, как и раньше, через ξ и ξ' расстояния предмета P и его изображения P' от центра линзы. С учетом правила знаков $u_1=-r/\xi$, $u_2=-r/\xi'$, так что предыдущее уравнение переходит в

$$\frac{f}{\xi}+\frac{f'}{\xi'}=-1, \tag{25.17}$$

где

$$\frac{1}{f} = -\frac{1}{n_1}\int\limits_{-\infty}^{+\infty}\left(\frac{\partial^2 n}{\partial r^2}\right)_0 dx, \qquad \frac{1}{f'} = \frac{1}{n_2}\int\limits_{-\infty}^{+\infty}\left(\frac{\partial^2 n}{\partial r^2}\right)_0 dx. \tag{25.18}$$

2. Вычислить фокусные расстояния тонкой электростатической линзы (примером такой линзы может служить «одиночная линза», представленная на рис. 110).

Решение. Потенциал поля V определим по формуле ${}^1\!/_2 mv^2 = |e|V$. При таком определении V — существенно положительная величина.

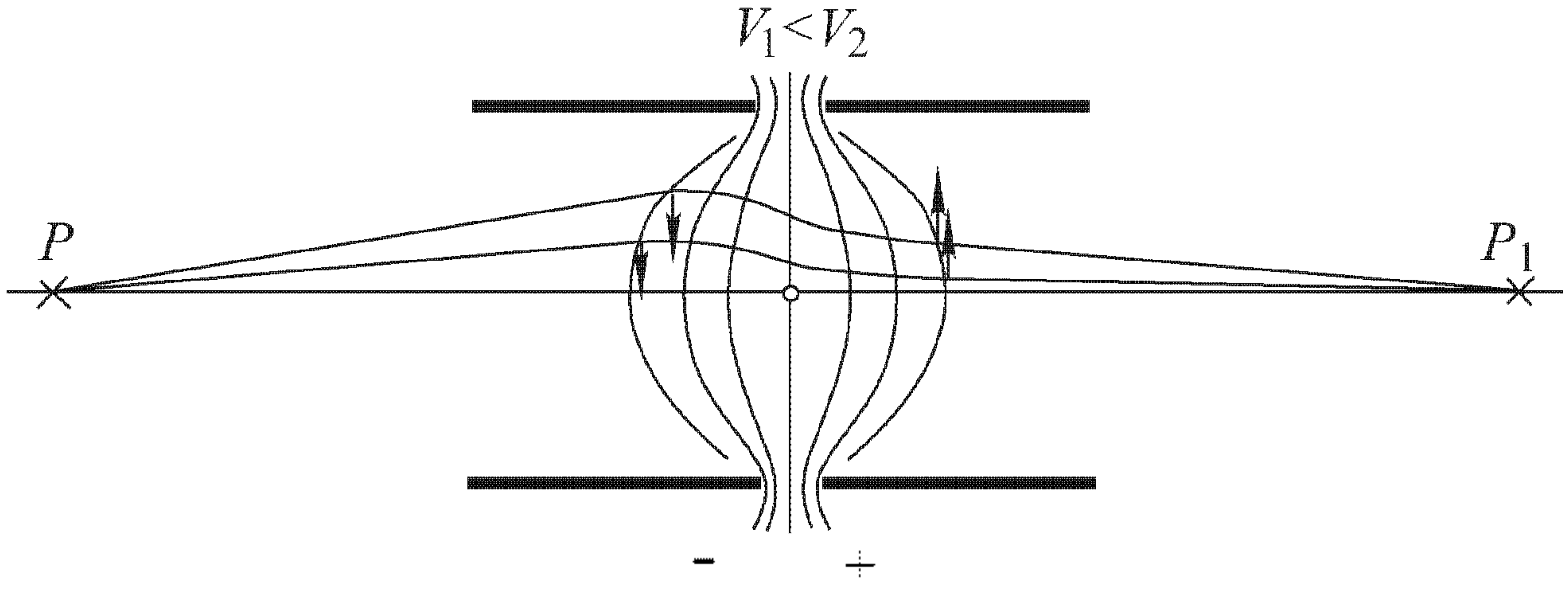


Рис. 110

В формулах (25.18) делаем замену $n \to \sqrt{V}$. Учитывая, что на оси системы $\partial V/\partial r = 0$, получаем

$$\left(\frac{\partial^2 n}{\partial r^2}\right)_0 \to \frac{\partial^2 \sqrt{V}}{\partial r^2} = \frac{1}{2\sqrt{V}}\frac{\partial^2 V}{\partial r^2}.$$

При отсутствии свободных зарядов $\nabla^2 V = 0$. На оси $\partial^2 V/\partial y^2 = \partial^2 V/\partial z^2 = \partial^2 V/\partial r^2$, так что $\nabla^2 V = \partial^2 V/\partial x^2 + 2(\partial^2 V/\partial r^2) = 0$. В результате из (25.18) находим

$$\frac{1}{f} = \frac{1}{4\sqrt{V_1}}\int\limits_{-\infty}^{+\infty}\frac{1}{\sqrt{V}}\frac{\partial^2 V}{\partial x^2}\,dx. \tag{25.19}$$

Проинтегрировав по частям и приняв во внимание, что на пределах интеграла $\partial V/\partial x = 0$, получим

$$\frac{1}{f} = \frac{1}{8\sqrt{V_1}}\int\limits_{-\infty}^{+\infty} V^{-3/2}E_x^2\,dx, \qquad \frac{1}{f'} = -\frac{1}{8\sqrt{V_2}}\int\limits_{-\infty}^{+\infty} V^{-3/2}E_x^2\,dx, \tag{25.20}$$

где V_1 — потенциал пространства предметов, V_2 — потенциал пространства изображений, $E_x = -\partial V/\partial x$ — напряженность электрического поля на оси. Если $V_1 = V_2$, то $f = -f'$.

Так как $V > 0$, то для тонких электростатических линз $f > 0$, $f' < 0$. Такие линзы всегда будут собирательными. Оптические линзы, даже тонкие, могут быть и собирательными, и рассеивающими. Это различие связано с тем, что функция $n(r)$ может быть какой угодно, тогда как потенциал

$V(r)$ при отсутствии свободных зарядов должен удовлетворять уравнению Лапласа $\nabla^2 V = 0$. Толстые электростатические линзы могут быть и собирательными, и рассеивающими. В этом нетрудно убедиться на примере системы, состоящей из двух тонких электростатических линз. Формулы (12.3) показывают, что такая система будет собирательной, когда оптический интервал Δ отрицателен, и рассеивающей, когда он положителен.

3. Предполагая, что $n_0 \gg \left|\frac{eA}{er}\right|$, вычислить фокусное расстояние тонкой магнитной линзы.

Решение. Разлагая (25.10) в ряд и отбрасывая члены высших степеней, получим

$$n = n_0 - \frac{e^2A^2}{2n_0c^2r^2} = n_0 - \frac{e^2r^2}{8n_0c^2}B_x^2,$$

откуда

$$\frac{\partial^2 n}{\partial r^2} = -\frac{e^2}{4n_0c^2}B_x^2.$$

Следовательно,

$$\frac{1}{f} = -\frac{1}{f'} = \frac{e^2}{4n_0^2c^2}\int_{-\infty}^{+\infty} B_x^2\,dx = \frac{e^2}{8mc^2W}\int_{-\infty}^{+\infty} B_x^2\,dx. \tag{25.21}$$

Как и электростатическая, тонкая магнитная линза всегда будет собирательной.

4. Вычислить фокусное расстояние кругового витка радиуса a тонкой проволоки, по которой течет ток I.

Ответ. В гауссовой системе единиц

$$\frac{1}{f} = \frac{e^2\pi^2a^4I^2}{8mc^4W}\int_{-\infty}^{+\infty}\frac{dx}{(x^2+a^2)^3} = \frac{3\pi^3e^2}{16mc^4W}\frac{I^2}{a}. \tag{25.22}$$

Выражая энергию W в электрон-вольтах, полагаем $W = |e|V_0$ и переходим к практическим единицам (сантиметры, вольты, амперы). Тогда получаем для электронов

$$f = -f' = 98\frac{V_0a}{I^2}. \tag{25.23}$$

Для короткой катушки с числом витков N

$$f = -f' = 98\frac{V_0a}{N^2I^2}. \tag{25.24}$$

5. Показать, что формула (25.18) для f в предельном случае тонкой линзы с резкими краями переходит в формулу (10.9). Считать, что показатели преломления n_1 и n_2 по обе стороны линзы одинаковы.

Указание. Принять, что показатель преломления n меняется только в бесконечно тонких слоях вблизи каждой сферической границы линзы. На всякой сферической поверхности, проведенной внутри этих слоев параллельно ближайшей поверхности линзы, функция n постоянна. Пользуясь этим, дифференцирование по r можно заменить дифференцированием по x, а затем вычислить интеграл в формулах (25.18).

ГЛАВА III

ИНТЕРФЕРЕНЦИЯ СВЕТА

§ 26. Общие сведения об интерференции

1. В *интерференции* и *дифракции* проявляются *волновые свойства света.* После открытия этих явлений на них смотрели сначала как на доказательство исключительно *волновой природы света.* Такая точка зрения оказалась недостаточной. В XX веке были открыты *корпускулярные свойства света*, а затем *волновые свойства обыкновенных частиц*: электронов, протонов, нейтронов, атомов, молекул и пр. Как это ни парадоксально, природа света и вещества оказалась *двойственной*: корпускулярно-волновой. С открытием этого факта связан коренной пересмотр физических воззрений, приведший к построению *квантовой механики.* От этого, однако, значение интерференционных и дифракционных явлений не уменьшилось. В наши дни интерференция и дифракция света имеют важные практические применения, например в спектроскопии и метрологии.

2. С помощью различных приспособлений, например отражений, преломлений или как-либо иначе, можно наложить один световой пучок на другой. Если выполняется принцип суперпозиции (а это предполагается всюду в настоящей главе), то каждый пучок будет проходить через область перекрытия так, как если бы другого пучка не было совсем. Пусть $\mathbf{E}_1$ — напряженность электрического поля, создаваемая первым пучком в произвольной точке A области перекрытия, а $\mathbf{E}_2$ — вторым. Согласно принципу суперпозиции, результирующая напряженность поля в той же точке A, создаваемая обоими пучками, будет представляться векторной суммой $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$. Так же ведут себя любые скалярные или векторные волны, какова бы ни была их физическая природа. Требуется только, чтобы они подчинялись *принципу суперпозиции.* Однако частота световых волн совсем иного порядка, чем частота, скажем, звуковых волн или волн на поверхности воды. Это ограничивает возможности приемников света.

Всем приемникам света присуща определенная *инерционность*. Ее можно характеризовать временем *установления* или *разрешения* приемника τ. Так, для глаза $\tau \approx 0{,}1$ с. Это есть время, в течение которого глаз сохраняет зрительное впечатление, т.е. «видит свет» даже тогда, когда излучение уже перестало в него попадать. Глаз не замечает, например, быстрых миганий света, если они следуют друг за другом через времена, малые по сравнению с τ (кино, телевидение). У фотоматериалов τ (время экспозиции) обычно порядка 10^{-2}–10^{-4} с. Существуют приемники, для которых время разрешения может быть сделано гораздо меньше. В ячейках Керра это время может быть доведено до 10^{-8}–10^{-9}. Наиболее быстродействующие современные фотоэлектрические приемники имеют время разрешения порядка 10^{-10} с, но оно может быть сделано еще меньше.

Однако даже такие времена очень велики по сравнению с периодами оптических колебаний. Средний период колебаний T электромагнитного поля в видимой области спектра составляет около 10^{-15} с. Поэтому ни один приемник света не позволяет измерить мгновенное значение напряженности электрического или магнитного поля в световой волне. Для этого время разрешения приемника должно было бы быть мало по сравнению с периодом световых колебаний T. Все приемники могут измерять только величины, *квадратичные по полю*, усредненные за времена, не меньшие времени разрешения приемника. Сюда относятся энергетические и фотометрические величины: лучистый и световой поток, яркость, освещенность и пр.

В явлениях интерференции, дифракции и пр. представляют интерес не абсолютные, а только *относительные значения* этих величин. Например, нас может интересовать относительное распределение освещенности на экране, куда попадает свет. При такой постановке вопроса нет смысла точно указывать, о какой именно энергетической или фотометрической величине идет речь в том или ином конкретном случае. Заключения будут относиться к любой усредненной по времени величине, квадратичной по напряженности электрического поля. Эту нечетко определенную величину принято называть *интенсивностью света* или *интенсивностью колебаний*. Ниже она обозначается через I. За I мы будем обычно принимать усредненное по времени значение квадрата напряженности электрического поля, т.е. $I = \overline{\mathbf{E}^2}$.

3. Найдем теперь интенсивность света в какой-либо точке пространства, где перекрываются два пучка света. Возведя равенство $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$ в квадрат и произведя усреднение по времени, получим

$$I = I_1 + I_2 + I_{12}, \tag{26.1}$$

где I_1 — интенсивность света первого пучка, а I_2 — второго. Последнее слагаемое

$$I_{12} = 2(\overline{\mathbf{E}_1\mathbf{E}_2}), \tag{26.2}$$

учитывающее взаимодействие пучков, называется *интерференционным членом.* Если взять независимые источники света, например две электрические лампочки, то повседневный опыт показывает, что $I = I_1 + I_2$, т.е. результирующая интенсивность равна сумме интенсивностей налагающихся пучков, а потому интерференционный член обращается в нуль. Тогда говорят, что пучки *не коррелированы* или *не когерентны* между собой. Однако, если накладывающиеся пучки не независимы, например один получается отражением другого от зеркала, то в некоторых случаях интерференционный член I_{12} не обращается в нуль, а потому $I \neq I_1 + I_2$. В одних точках пространства результирующая интенсивность I больше, в других меньше суммы интенсивностей I_1 и I_2. Это явление называется *интерференцией волн*, или *интерференцией колебаний.* Говорят, что источники света и соответствующие им пучки, дающие интерференцию, *коррелированы* или *когерентны* между собой, т.е. колебательные и волновые процессы в них протекают во времени согласованно. Говорят также, что когерентные пучки *интерферируют*, а некогерентные *не интерферируют* между собой.

Все сказанное о двух пучках относится и к случаю наложения нескольких пучков. Интерференция двух пучков называется *двухлучевой*, многих пучков — *многолучевой.*

4. Начнем с идеализированного случая, когда обе волны *строго монохроматические* и имеют одну и ту же частоту. *Монохроматическая волна* — это строго синусоидальная волна с постоянными во времени частотой ω, амплитудой a и начальной фазой φ. Амплитуда и фаза колебаний могут меняться от одной точки пространства к другой, частота одна и та же для колебательного процесса *во всем пространстве.* Монохроматическое колебание в каждой точке пространства длится *бесконечно долго*, не имея ни начала, ни конца во времени. Поэтому строго монохроматические колебания и волны никогда не могут быть точно реализованы в действительности. Однако эти идеализации играют громадную роль в учении о колебаниях и волнах, в чем мы уже убедились в гл. X третьего тома и еще в большей степени убедимся в дальнейшем.

Допустим сначала, что в рассматриваемой точке наблюдения оба вектора $\mathbf{E}_1$ и $\mathbf{E}_2$ параллельны или антипараллельны. Тогда можно отвлечься от векторного характера колебаний, считая их скалярными. Представим эти колебания в вещественной форме:

$$E_1 = a_1 \cos(\omega t + \varphi_1), \quad E_2 = a_2 \cos(\omega t + \varphi_2), \tag{26.3}$$

где a_1 и a_2, φ_1 и φ_2 — амплитуды и начальные фазы колебаний. Если ввести комплексные амплитуды $A_1 = a_1 e^{i\varphi_1}$ и $A_2 = a_2 e^{i\varphi_2}$, то в комплексной форме те же колебания представятся так:

$$E_1 = A_1 e^{i\omega t}, \quad E_2 = A_2 e^{i\omega t}. \tag{26.4}$$

Результирующее колебание будет

$$E = E_1 + E_2 = (A_1 + A_2) e^{i\omega t}.$$

Это — также монохроматическое колебание с той же частотой ω и комплексной амплитудой $A = A_1 + A_2$. Чтобы найти вещественную амплитуду a и начальную фазу φ результирующего колебания, запишем последнее соотношение так:

$$a e^{i\varphi} = a_1 e^{i\varphi_1} + a_2 e^{i\varphi_2}.$$

Умножая его на комплексно-сопряженное, получим

$$a^2 = a_1^2 + a_2^2 + 2a_1 a_2 \cos(\varphi_2 - \varphi_1), \tag{26.5}$$

а после отделения вещественной части от мнимой

$$a \cos\varphi = a_1 \cos\varphi_1 + a_2 \cos\varphi_2, \quad a \sin\varphi = a_1 \sin\varphi_1 + a_2 \sin\varphi_2.$$

Отсюда

$$\operatorname{tg}\varphi = \frac{a_1 \sin\varphi_1 + a_2 \sin\varphi_2}{a_1 \cos\varphi_1 + a_2 \cos\varphi_2}. \tag{26.6}$$

На рис. 111 приведена *векторная диаграмма* сложения рассматриваемых колебаний, из которой также нетрудно получить результаты (26.5) и (26.6).

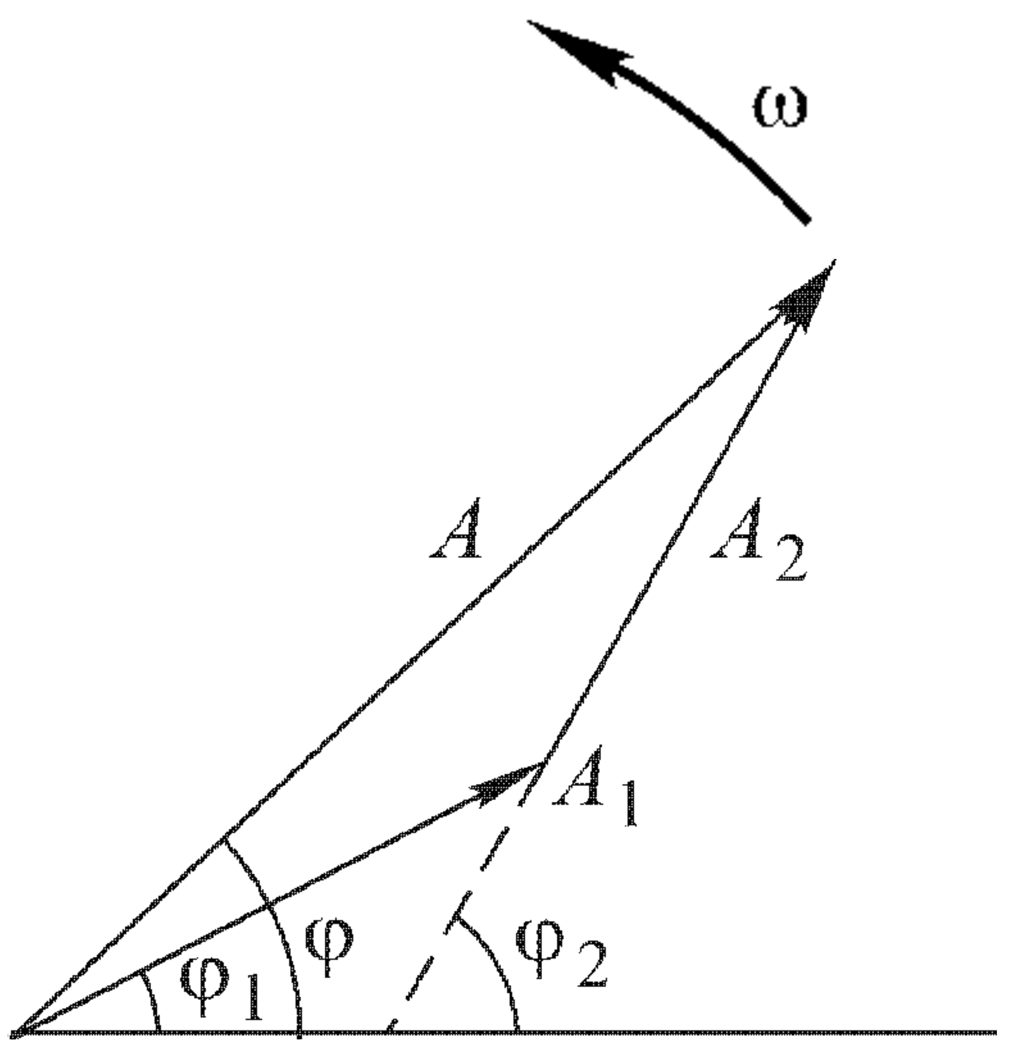


Рис. 111

Вводя интенсивности колебаний, результат (26.5) можно записать в виде

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\varphi_2 - \varphi_1). \tag{26.7}$$

Если колебания синфазны, т.е. фазы φ_1 и φ_2 одинаковы или отличаются на четное число π, то интенсивность I максимальна и равна

$$I_{\text{макс}} = \left(\sqrt{I_1} + \sqrt{I_2}\right)^2. \tag{26.8}$$

Если колебания противофазны, т.е. фазы φ_1 и φ_2 отличаются на нечетное число π, то получается минимальная интенсивность:

$$I_{\text{мин}} = \left(\sqrt{I_1} - \sqrt{I_2}\right)^2. \tag{26.9}$$

Если колебания совершаются в квадратуре, т.е. их фазы отличаются на $m\pi \pm \pi/2$ (m — целое число), то $I = I_1 + I_2$. В этом случае интенсивность результирующего колебания равна сумме интенсивностей складываемых колебаний.

5. Не представляет труда написать интерференционный член и для общего случая, когда складываются *векторные* колебания, причем между декартовыми компонентами каждого вектора могут существовать произвольные разности фаз. Предоставляя это сделать читателю, заметим, что Френель и Араго обнаружили на опыте, что *две световые волны, распространяющиеся в одном направлении, никогда не интерферируют между собой, если они линейно поляризованы во взаимно перпендикулярных плоскостях.* На этом основании Френель пришел к заключению, что световые волны *поперечны.*

Покажем это, не вводя никаких специальных предположений о физической природе «светового вектора» $\mathbf{E}$, совершающего колебания в световой волне. Допустим, что вдоль оси Z распространяются две волны, плоскости колебаний которых взаимно перпендикулярны. В одной волне колебания происходят в координатной плоскости XZ, в другой — в координатной плоскости YZ. Представим световые векторы этих волн в виде $\mathbf{E}_1 = \mathbf{E}_{1x} + \mathbf{E}_{1z}$ и $\mathbf{E}_2 = \mathbf{E}_{2y} + \mathbf{E}_{2z}$, где координатные индексы указывают, какой из координатных осей параллелен соответствующий вектор. Перемножая скалярно и усредняя по времени, находим интерференционный член:

$$I_{12} = 2(\overline{\mathbf{E}_{1z}\mathbf{E}_{2z}}) = 2a_{1z}a_{2z}\cos(\varphi_{2z} - \varphi_{1z}),$$

где a_{1z}, a_{2z} — амплитуды, а φ_{1z}, φ_{2z} — фазы соответствующих продольных колебаний. Опыты Френеля и Араго показали, что интерференционный член обращается в нуль, каковы бы ни были фазы колебаний. Отсюда следует, что по крайней мере одна из амплитуд, например a_{1z}, равна нулю, т.е. первая волна поперечна. Но тогда и вторая волна должна быть поперечной, так как нет никаких оснований предпочесть первую волну второй.

6. Допустим теперь, что перекрываются две плоские волны:

$$E_1 = a_1\cos(\omega t - \mathbf{k}_1\mathbf{r} + \delta_1), \quad E_2 = a_2\cos(\omega t - \mathbf{k}_2\mathbf{r} + \delta_2). \tag{26.10}$$

Снова предположим, что векторы $\mathbf{E}_1$ и $\mathbf{E}_2$ параллельны или антипараллельны, так что от векторного характера колебаний можно отвлечься. Сравнивая эти выражения с (26.3), видим, что в рассматриваемом случае

$$\varphi_2 - \varphi_1 = \mathbf{K}\mathbf{r} + (\delta_2 - \delta_1), \tag{26.11}$$

где введен новый вектор $\mathbf{K} = \mathbf{k}_1 - \mathbf{k}_2$. Он параллелен биссектрисе угла, внешнего по отношению к углу α, который составляют между собой волновые векторы $\mathbf{k}_1$ и $\mathbf{k}_2$ (рис. 112). Поверхности равных разностей фаз $\varphi_2 - \varphi_1 = \text{const}$ суть параллельные плоскости, перпендикулярные к вектору $\mathbf{K}$. Они обозначены на рис. 112 штриховыми прямыми. Вдоль каждой из этих плоскостей, следовательно, интенсивность результирующего колебания будет постоянна. Она максимальна, когда разность фаз $\varphi_2 - \varphi_1$ содержит π четное число раз, и минимальна, когда π содержится нечетное число раз. В частности, если складываются волны с одной и той же интенсивностью I, то интенсивность в максимуме будет $4I$, а в минимуме — нуль. Расстояние Δx между двумя соседними плоскостями максимальной или минимальной интенсивности найдется из условия $K\,\Delta x = 2\pi$. Так как длины обоих волновых векторов $\mathbf{k}_1$ и $\mathbf{k}_2$ одинаковы и равны $k = 2\pi/\lambda$, то $K = 2k\sin(\alpha/2)$, и следовательно,

Рис. 112

$$\Delta x = \frac{2\pi}{K} = \frac{\pi}{k\sin(\alpha/2)} = \frac{\lambda}{2\sin(\alpha/2)}. \tag{26.12}$$

Для малых углов α

$$\Delta x \approx \lambda/\alpha. \tag{26.13}$$

Если поставить плоский экран, то он пересечет плоскости равной интенсивности вдоль параллельных прямых; на экране появятся светлые и темные «интерференционные полосы». Расстояние между серединами соседних светлых или темных полос называется *шириной интерференционной полосы*. Если плоскость экрана параллельна плоскости $(\mathbf{k}_1, \mathbf{k}_2)$, в которой лежат волновые векторы $\mathbf{k}_1$ и $\mathbf{k}_2$, то ширина интерференционной полосы равна Δx, т.е. определяется выражением (26.12). То же самое получится, если экран установлен в перпендикулярной плоскости перпендикулярно к биссектрисе угла между волновыми векторами $\mathbf{k}_1$ и $\mathbf{k}_2$. Если же, оставляя экран перпендикулярным к плоскости $(\mathbf{k}_1, \mathbf{k}_2)$, повернуть его на угол φ, то ширина интерференционной полосы сделается равной $\Delta_\varphi x = \Delta x/\cos\varphi$.

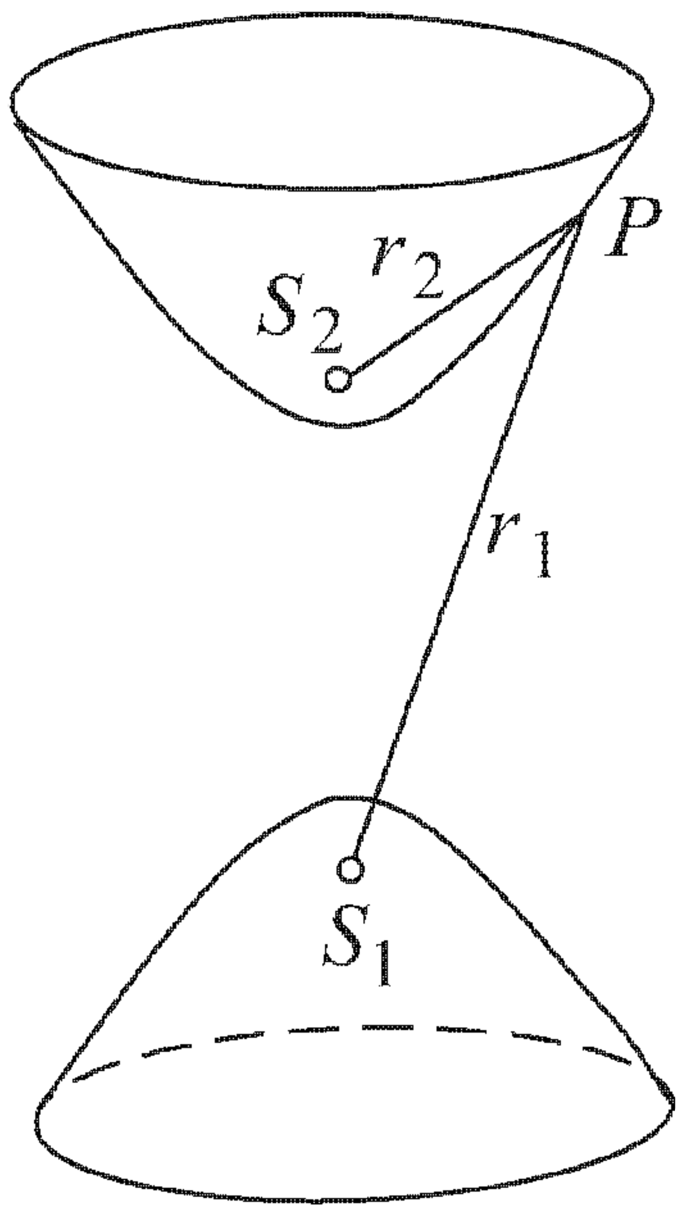


Рис. 113

7. Разберем теперь случай, когда перекрываются *сферические монохроматические волны* от двух точечных источников света S_1 и S_2 (рис. 113). В этом случае амплитуды a_1 и a_2 складываемых колебаний обратно пропорциональны расстояниям r_1 и r_2 от точки наблюдения до источников S_1 и S_2. Поэтому интенсивность света будет меняться вдоль каждой интерференционной полосы. Однако это изменение медленное, и от него можно отвлечься. Основное значение для результирующей интенсивности имеет разность фаз складываемых колебаний. Поверхности равных разностей фаз $r_1 - r_2 = \mathrm{const}$ будут двухполостными гиперболоидами вращения с фокусами S_1 и S_2. На экране, перпендикулярном к линии источников S_1S_2, получатся интерференционные полосы в виде концентрических колец с центром в точке пересечения экрана с указанной линией. Если же экран параллелен линии источников S_1S_2, то интерференционные полосы будут гиперболами с фокусами S_1 и S_2. В последнем случае в небольшой центральной части интерференционная картина практически будет состоять из равноотстоящих параллельных светлых и темных полос.

Если фазы колебаний обоих источников света одинаковы, то разность фаз складываемых колебаний будет

$$\Delta\varphi = k(r_2 - r_1) = (2\pi/\lambda)(r_2 - r_1).$$

Когда она равна $2m\pi$ (m — целое число, положительное или отрицательное), то получается максимум интенсивности колебаний (светлая интерференционная полоса). Если же $\Delta\varphi = 2\pi(m+{}^1\!/_2)$, то интерференционная полоса будет темной. Условия максимума и минимума можно также записать в виде

$$\Delta r \equiv r_2 - r_1 = \begin{cases} m\lambda & \text{(светлая полоса)}, \\ (m + {}^1\!/_2)\lambda & \text{(темная полоса)}. \end{cases} \tag{26.14}$$

Величина $r_2 - r_1$ называется *разностью хода* интерферирующих лучей. Если интерферирующие лучи проходят через среды с различными показателями преломления, то величину $r_2 - r_1$ надо заменить на

$$\Delta = \int n_2\,dl - \int n_1\,dl,$$

т.е. на *оптическую разность хода* интерферирующих лучей. Для светлых интерференционных полос оптическая разность хода составляет целое, для темных — полуцелое число длин волн (в вакууме). Целое число m называется *порядком интерференции*. Порядок интерференции есть округленная до целого числа оптическая разность хода интерферирующих лучей, выраженная в длинах волн (в вакууме).

8. Основные результаты, найденные выше, можно также получить на следующем простом примере. Пусть S_1 и S_2 — близко расположенные точечные монохроматические источники света (рис. 114). Для увеличения интенсивности интерференционных полос вместо них можно взять два коротких линейных источника, например две узкие ярко освещенные щели. Плоскость экрана *Э* предполагается параллельной плоскости, в которой лежат линейные источники S_1 и S_2. Пусть CO — перпендикуляр к этой плоскости, проходящий через середину между источниками S_1

и S_2, D — длина этого перпендикуляра, d — расстояние между источниками S_1 и S_2. Предполагается, что не только расстояние d, но также длины источников и линейные размеры экрана малы по сравнению с расстоянием D. Тогда интерференционные полосы на экране будут прямолинейны и перпендикулярны к линии, соединяющей источники S_1 и S_2. Начало координат поместим в точке O на экране, ось X направим параллельно линии источников S_1S_2. Если x — абсцисса точки наблюдения A, то

$$r_1^2 = D^2 + (x + d/2)^2, \quad r_2^2 = D^2 + (x - d/2)^2,$$

так что $r_1^2 - r_2^2 = 2xd$, и следовательно, $r_1 - r_2 = 2xd/(r_1 + r_2)$. Так как $x \ll D$, то без существенной ошибки знаменатель $r_1 + r_2$ можно заменить на $2D$. Тогда

$$r_1 - r_2 = \frac{xd}{D} = \alpha x \; ^{1)}, \tag{26.15}$$

где $\alpha \approx d/D$ — угол схождения интерферирующих лучей, т.е. угол, под которым из точки O (или в рассматриваемом приближении из любой точки экрана) видно расстояние d между источниками S_1 и S_2. Интенсивность света найдется по формулам (26.7) и (26.11). Если источники S_1 и S_2 одинаковы и синфазны, то эти формулы дают

$$I = 2I_1\left(1 + \cos\frac{2\pi\alpha x}{\lambda}\right). \tag{26.16}$$

Интенсивность I периодически меняется вдоль оси X от нуля до максимального значения $I_{\text{макс}} = 4I_1$. Пространственный период изменения интенсивности $\Delta x = \lambda/\alpha$ есть ширина интерференционной полосы.

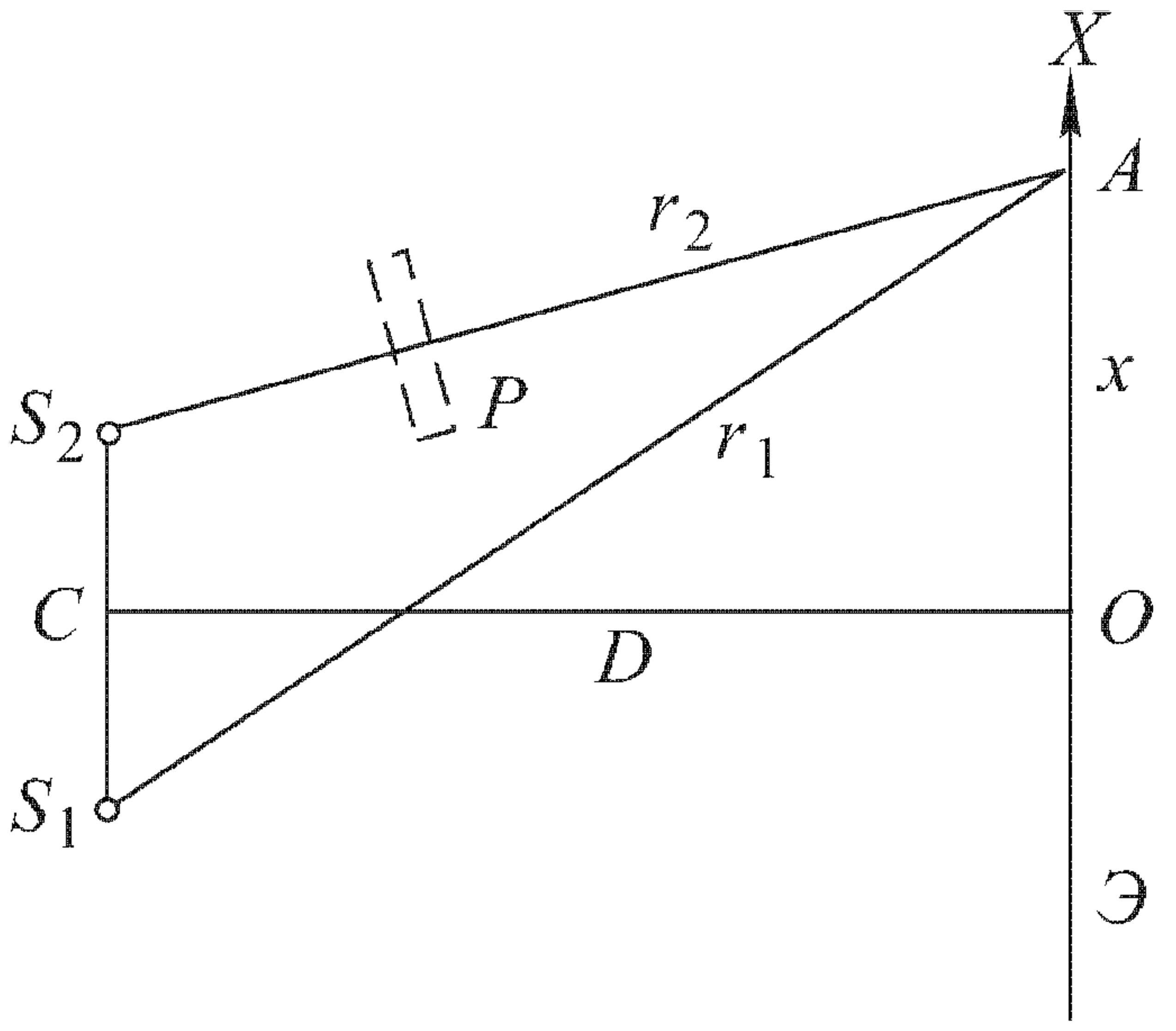


Рис. 114

Если на пути одного из лучей, например S_2A, ввести прозрачную плоскопараллельную пластинку P с толщиной l и показателем преломления n, то оптическая длина этого луча увеличится на $(n-1)l$, а разность хода между лучами S_1A и S_2A уменьшится на такую же величину. Прежнее значение разности хода получится в какой-то другой точке A', отстоящей от S_1 и S_2 на

---

[1] Если пренебречь членами третьей степени по d, то $r_1 - r_2 = xd/\sqrt{D^2 + x^2}$. Пренебрегая далее пятыми степенями x/D, отсюда получим

$$r_1 - r_2 = \frac{xd}{D}\left(1 - \frac{x^2}{2D^2}\right) \approx \alpha x\left(1 - \frac{x^2}{2D^2}\right). \tag{26.15а}$$

расстояниях r_1' и r_2'. Положение точки A' найдется из условия

$$r_1' - [r_2' + (n-1)l] = r_1 - r_2, \quad \text{или} \quad (r_1' - r_2') = (r_1 - r_2) + (n-1)l.$$

Это значит, что произойдет смещение всей интерференционной картины на $N = (n-1)l/\Delta x$ полос в ту сторону, с какой была введена пластинка P. На этом основаны интерференционные методы измерения малых изменений показателя преломления, обладающие высокой чувствительностью. Для таких измерений надо пользоваться белым светом (см. § 30). В монохроматическом свете все полосы одинаковы, и их смещение трудно измерить — для этого толщину пластинки l надо было бы увеличивать непрерывно от нуля до окончательной величины.

9. Наличие интерференционных максимумов и минимумов в волновом поле, конечно, не может привести к противоречию с принципом сохранения энергии. Обсудим этот вопрос для только что исследованного случая двух одинаковых синфазных монохроматических точечных источников S_1 и S_2.

Максимумы интенсивности находятся на гиперболах $r_1 - r_2 = m\lambda$, минимумы — на гиперболах $r_1 - r_2 = (m + {}^1\!/_2)\lambda$. Максимальное значение, которое может принимать разность $r_1 - r_2$, равно расстоянию между источниками d, а максимальное значение m — целой части, содержащейся в числе d/λ. Когда $d \gg \lambda$, число интерференционных полос в волновом поле велико. В этом случае интерференция приводит только к *пространственному перераспределению* лучистой энергии, в результате чего в одних местах пространства (в максимумах) плотность лучистой энергии увеличивается, в других (в минимумах) уменьшается, а общий запас энергии в пространстве остается неизменным.

Полный поток лучистой энергии через любую замкнутую поверхность, окружающую источники S_1 и S_2, равен сумме потоков через ту же поверхность, которые излучались бы каждым из этих источников в отсутствие другого. Однако это равенство *не точное, а только приближенное* и в среднем выполняется тем точнее, чем больше d/λ. Когда $d < \lambda$, в волновом поле уже нет линии нулевой интенсивности. Когда же $d \ll \lambda$, то во всех точках пространства складываются колебания, фазы которых практически одинаковы. В этом случае интенсивность результирующего колебания, а с ним и поток результирующего излучения, исходящий от обоих источников, в четыре раза больше соответствующих величин для одного из источников в отсутствие другого. Таким образом, если расстояние между источниками S_1 и S_2 меньше длины волны λ, то поток лучистой энергии через замкнутую поверхность, окружающую эти источники, *больше суммы потоков*, которые излучали бы те же, но уединенные источники.

Однако принцип сохранения энергии и не требует равенства этих величин. Источники вместе действительно больше излучают энергии, чем в том случае, когда они находятся далеко друг от друга. Но это увеличение излучения происходит не за счет нарушения сохранения энергии, а *за счет работы генератора*, который должен поддерживать постоянными амплитуды колебаний в источниках. Если же колебания в источниках свободные, то увеличение излучения приводит просто к более быстрому затуханию этих колебаний.

Рассмотренный случай трудно реализовать в оптическом диапазоне спектра из-за малости длин световых волн. Но в области радиоволн это не только возможно, но и практически используется для получения *направленных излучений и увеличения их мощности.* Конечно, возможен и такой случай, когда совместное действие двух источников приводит не к увеличению, а *к уменьшению* общего излучения. Это будет, например, когда колебания в источниках происходят в противоположных фазах.

10. Строго монохроматические волны одной и той же частоты, даже от двух различных точечных источников, *всегда интерферируют между собой*, т.е. когерентны. Интерференционная картина от таких источников *устойчива* в том смысле, что распределение интенсивности волнового поля в пространстве остается *неизменным во времени.* Интерференция излучений от *независимых источников света*, например от электрических лампочек, недоступна наблюдению глазом. Глаз не наблюдает интерференцию даже при наложении излучений от двух независимых самых узких спектральных линий, излучаемых разреженными газами. Это служит лучшим доказательством того, что излучения реальных источников света никогда не бывают строго монохроматичными.

Невозможность визуального наблюдения интерференционных полос от независимых источников света можно пояснить на примере идеализированных источников, излучающих *квазимонохроматический свет.* Такой свет представляется колебаниями вида (26.3), в которых, однако, амплитуды a_1, a_2 и фазы φ_1, φ_2 медленно и хаотически меняются во времени, т.е. испытывают заметные изменения за времена, очень большие по сравнению с периодом T самих световых колебаний. Примером может служить излучение изолированного атома. Возбужденный атом испускает ряд или, как принято говорить, *цуг волн* в течение времени $\tau_{\text{изл}}$, характерная длительность которого порядка 10^{-8} с (см. § 89). В таком цуге содержится 10^6–10^8 волн. За время $\tau_{\text{изл}}$ атом «высвечивается» и переходит в невозбужденное состояние. В результате различных процессов, например столкновений с другими атомами или ударов электронов, атом может снова вернуться в возбужденное состояние, а затем начать излучать новый цуг волн. Таким образом, получится последовательность цугов, испускаемых атомом через малые и нерегулярно меняющиеся промежутки времени. Пусть теперь на экран попадают излучения от двух независимых атомов. При наложении двух цугов, излучаемых этими атомами, на экране получится какая-то картина интерференционных полос. Положение полос определяется разностью фаз между колебаниями обоих цугов. А такая разность фаз быстро и беспорядочно меняется от одной пары цугов к следующей. В течение секунды десятки и сотни миллионов раз или чаще одна система интерференционных полос будет сменяться другой. Глаз или другой приемник света не в состоянии следить за этой быстрой сменой интерференционных картин и фиксирует только равномерную освещенность экрана.

Смещение интерференционных полос при замене одной пары цугов другой происходить не будет, если у этих цугов начальные фазы δ_1 и δ_2 хотя и меняются нерегулярно во времени, но *одинаково*. Действительно, в этом случае $\delta_2 - \delta_1 = 0$ и, как видно из (26.11), разность фаз складываемых колебаний $\varphi_2 - \varphi_1$ содержит только слагаемое $\mathbf{Kr}$, *регулярно* меняющееся в пространстве. Чтобы этого достигнуть, надо излучение от *одного и того же источника* расщепить на два или несколько пучков и заставить их попадать на экран *различными путями*. Если при этом будут накладываться пучки, полученные из *одного и того же цуга*, то интерференционные полосы на экране будут сохранять свое положение и глаз зафиксирует устойчивую (неподвижную) интерференционную картину. Некоторые опыты, основанные на таком принципе, будут описаны в следующем параграфе.

11. Отметим, что *нет принципиальных запретов* на возможность получения интерференции от двух *независимых* источников света. Такая возможность определяется *уровнем техники генерирования и приема излучения.* Еще в 1947 г. Г.С. Горелик (1906–1957) предлагал осуществить *гетеродинирование*, или *демодуляцию* видимого света. В видимом свете можно выделить дублеты или мультиплеты, т.е. сложные спектральные линии, состоящие из двух или нескольких одиночных спектральных линий с разностями частот $\nu_1 - \nu_2$ порядка 10^9–10^{10} Гц. По идее Горелика такой свет должен падать на катод фотоэлемента (а еще лучше фотоумножителя). Если верно предположение, что мгновенный фототок пропорционален квадрату напряженности результирующего электрического поля, то его можно представить в виде

$$\mathcal{I} = A(\mathbf{E}_1 + \mathbf{E}_2)^2 = A(\mathbf{E}_1^2 + \mathbf{E}_2^2 + 2\mathbf{E}_1\mathbf{E}_2),$$

где $\mathbf{E}_1$ и $\mathbf{E}_2$ — напряженности электрического поля, соответствующие частотам ν_1 и ν_2, а A — постоянная прибора. В этом выражении слагаемое $2A(\mathbf{E}_1\mathbf{E}_2)$ содержит составляющую с разностной частотой $\nu_1 - \nu_2$. Частоте $\nu_1 - \nu_2 \sim 10^{10}$ Гц соответствует длина волны $\lambda \sim 3$ см, лежащая в микроволновой области радиоволн. Ее можно усилить радиотехническими средствами и сделать доступной наблюдению. Тем самым биения интенсивности, возникающие при наложении пучков света от двух источников, превратятся в колебания электрического тока.

Для осуществления опыта в принципе безразлично, испускаются ли волны с частотами ν_1 и ν_2 одним источником или двумя независимыми источниками света. Трудность опыта состоит в том, что на выделяемый сигнал разностной частоты $\nu_1 - \nu_2$ накладывается «шум», интенсивность которого превосходит интенсивность сигнала примерно в 10^4 раз. Эта трудность была преодолена в 1955 г. Форрестором и его сотрудниками. Источником света служила яркая одиночная зеленая линия ртути $\lambda = 546{,}1$ нм, расщеплявшаяся в магнитном поле (эффект Зеемана). Применялся фотоэлемент с сурьмяно-цезиевым катодом. Положительный результат опыта доказывает, что если и существует запаздывание выхода электронов с поверхности фотокатода по сравнению с попаданием на нее фотонов, то оно значительно меньше 10^{-10} с.

С изобретением лазеров техника эксперимента получила источники света столь высокой степени монохроматичности, что наблюдение интерференционных максимумов и минимумов от двух независимых лазеров с помощью надлежащих приемников света (например, тех же фотоэлементов) стало сравнительно легко осуществимой экспериментальной работой.

§ 27. Классические интерференционные опыты

1. Опыт Юнга. Юнг получал полосы интерференции по способу, описанному им в публичных лекциях 1807 г. Яркий пучок света от Солнца падал на экран с малым отверстием или узкой щелью S (рис. 115). Дифрагированный свет шел ко второму экрану с двумя узкими отверстиями или щелями S_1 и S_2. На этих щелях свет также претерпевал дифракцию, в результате чего получались два перекрывающихся расходящихся пучка света с вершинами в S_1 и S_2. Ввиду общности происхождения эти пучки когерентны. На экране в месте перекрытия пучков наблюдались параллельные интерференционные полосы (см. предыдущий параграф, п. 8). Расстояние между щелями S_1 и S_2 должно быть велико по сравнению с шириной каждой щели.

Удобный способ получения нужных щелей был указан Рэлеем. Стеклянные пластинки покрываются тонким слоем серебра и тем самым становятся непрозрачными. На серебряном слое одной из пластинок лезвием бритвы проводится одна, а на другой две близко расположенные параллельные линии, которые и используются в качестве щелей.

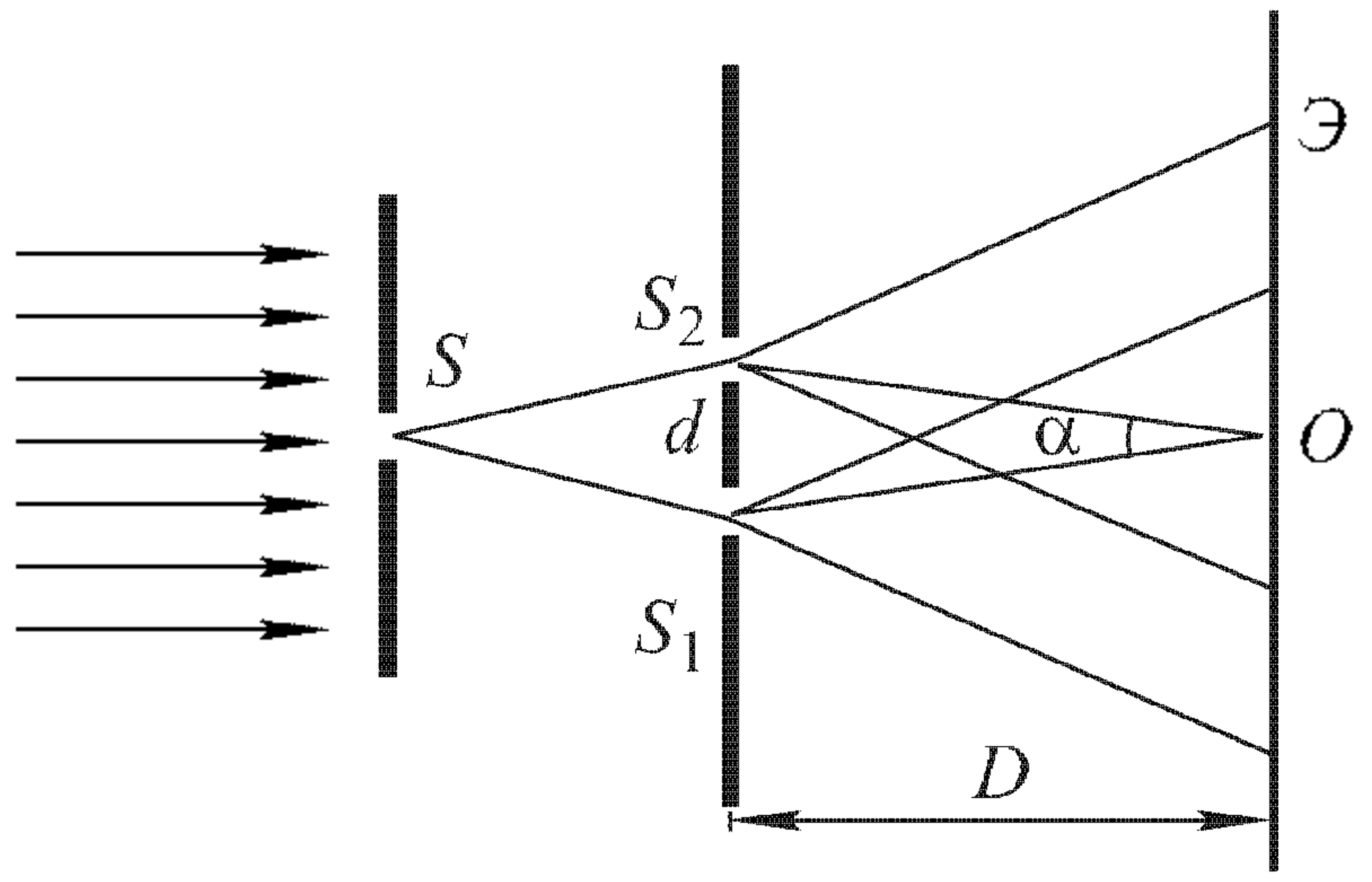


Рис. 115

Оценим по формуле (26.13) ширину Δx интерференционной полосы. Допустим, что расстояние d между щелями S_1 и S_2 составляет 1 мм, а расстояние от щелей до экрана $D = 1$ м. Тогда $\alpha = d/D = 0{,}001$ рад. Для красного света ($\lambda = 600$ нм) получаем $\Delta x = \lambda/\alpha = 6 \cdot 10^5$ нм $= 0{,}6$ мм. В синем свете ширина полосы будет порядка 0,4 мм. Таким путем впервые Юнг измерил длины световых волн, хотя эти измерения и не могли быть точными. В опыте Юнга из-за дифракции свет распространяется от щелей S_1 и S_2 в различных направлениях неравномерно. Влияние этого осложняющего обстоятельства будет выяснено в гл. IV.

Задолго до Юнга, в 1665 г., аналогичный опыт был поставлен Гримальди. Однако в опыте Гримальди свет от Солнца падал непосредственно на щели S_1 и S_2. Дополнительной щели S не было. При такой постановке опыта интерференционные полосы получаться не могли, ввиду значительных угловых размеров Солнца (см. § 28, п. 7).

При использовании лазеров, генерирующих практически параллельные пучки лучей, щель S в опыте Юнга не нужна.

2. Зеркала Френеля. В 1816 г. Френель осуществил следующий интерференционный опыт. Свет от узкой ярко освещенной щели S (рис. 116) падал на два плоских зеркала CD и CE, наклоненных друг к другу под углом, близким к 180°. Щель S устанавливалась параллельно линии C пересечения плоскостей зеркал. При отражении падающий пучок разделялся на два когерентных пучка, как бы исходивших от прямых S_1 и S_2, являющихся мнимыми изображениями щели S в зеркалах. Прямой свет от источника S загораживался непрозрачной ширмой MN. На экран Э попадали только отраженные пучки. В области AB, где пучки перекрывались, наблюдались параллельные интерференционные полосы. Для расчета ширины интерференционной полосы Δx введем следующие обозначения:

$CO = a$, $CS = b$, $\angle HCD = \varphi$. Так как точки S, S_1 и S_2 лежат на одной окружности радиуса b, то, предполагая угол φ малым, можем написать $d = S_1S_2 = 2b\varphi$. Угол α, под которым из точки O видно расстояние S_1S_2, равен $\alpha = d/(a+b) = 2b\varphi/(a+b)$, а потому

$$\Delta x = \frac{\lambda}{\alpha} = \frac{a+b}{2b\varphi}\lambda. \tag{27.1}$$

Угол α можно просто измерить по шкале зрительной трубы. Для этого трубу следует поместить в точке O и установить ее на отчетливое видение изображений S_1 и S_2 щели S. Тогда можно найти λ по формуле $\lambda = \alpha\,\Delta x$. Ширина области перекрытия $AB = 2a\varphi$, а потому число полос, которые могут наблюдаться на экране, равно

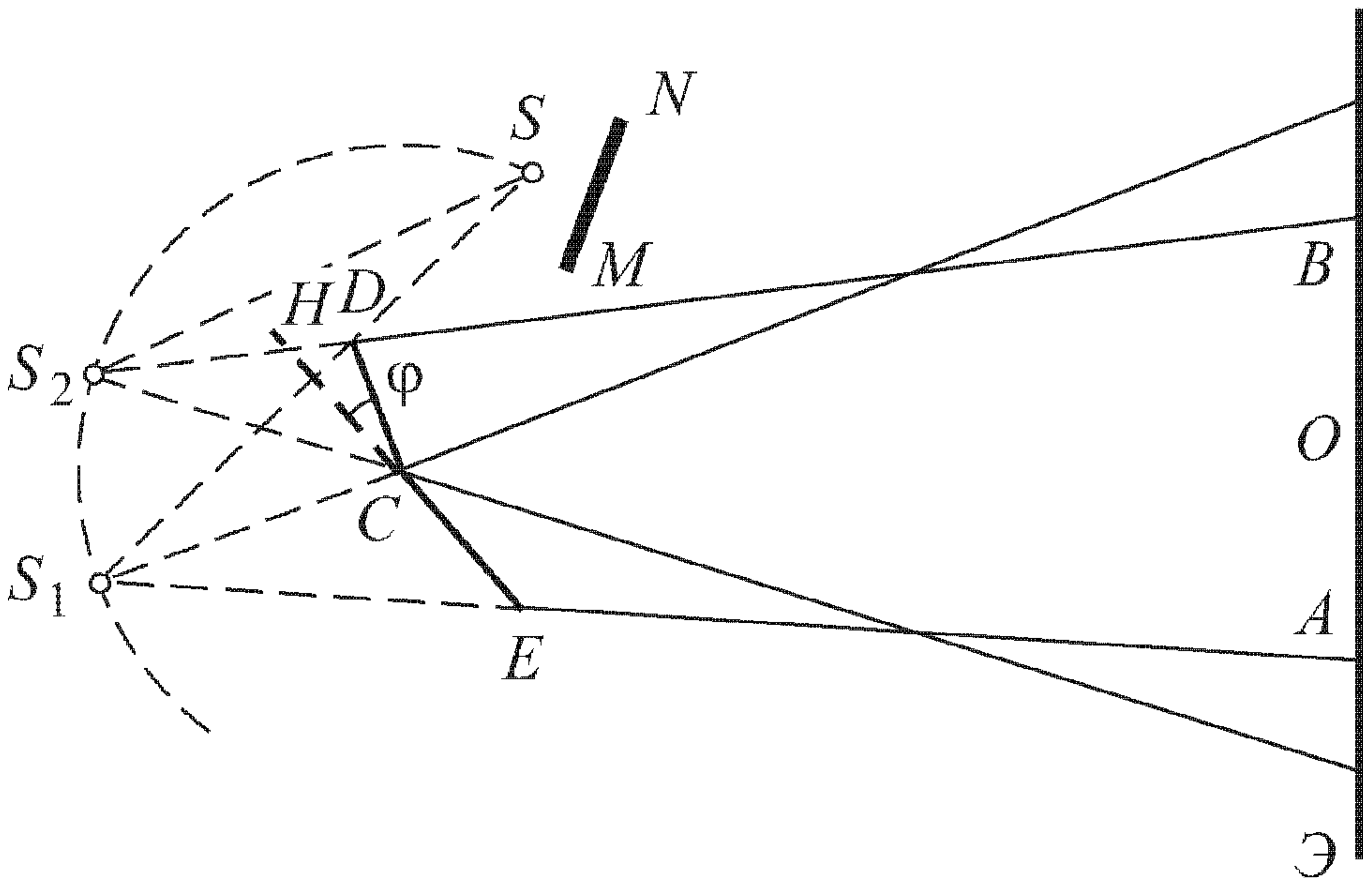


Рис. 116

$$N = \frac{2a\varphi}{\Delta x} = \frac{2a\varphi\alpha}{\lambda} = \frac{4ab}{\lambda(a+b)}\varphi^2. \tag{27.2}$$

В опыте Френеля интерференционная картина искажается дифракцией на ребре C, вдоль которого пересекаются плоскости зеркал.

3. Бипризма Френеля. Бипризма Френеля состоит из двух стеклянных призм с малыми преломляющими углами, сложенных своими основаниями (рис. 117). Практически она изготовляется из целого куска стекла. Источником света служит ярко освещенная щель S, устанавливаемая параллельно ребру бипризмы. После преломления в бипризме падающий пучок света разделяется на два когерентных пучка с вершинами в мнимых изображениях S_1 и S_2 щели S. В области AB экрана пучки перекрываются и дают систему параллельных интерференционных полос. Пусть β — преломляющий угол бипризмы, n — ее показатель преломления, a и b — длины отрезков CO и SC. Каждая половина бипризмы отклоняет параксиальный луч на угол $(n-1)\beta$. Расстояние d между изображениями S_1 и S_2 равно $d = S_1S_2 = 2b(n-1)\beta$, а угловое расстояние между ними $\alpha = d/(a+b)$. Ширина полосы

$$\Delta x = \frac{\lambda}{\alpha} = \frac{\lambda(a+b)}{2b(n-1)\beta}. \tag{27.3}$$

Искажения интерференционной картины, вносимые дифракцией на ребре C бипризмы, более существенны, чем в опыте с зеркалами Френеля. Однако осуществить опыт с бипризмой значительно легче.

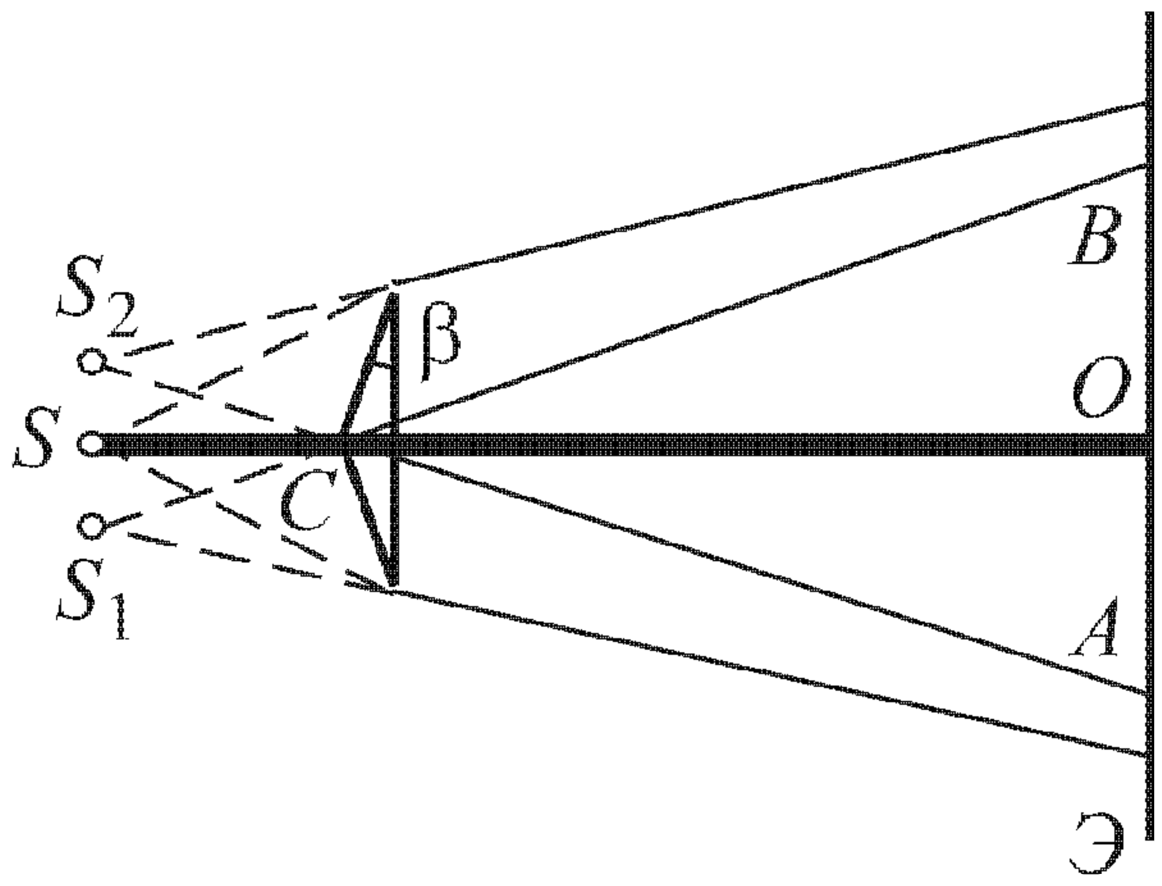


Рис. 117

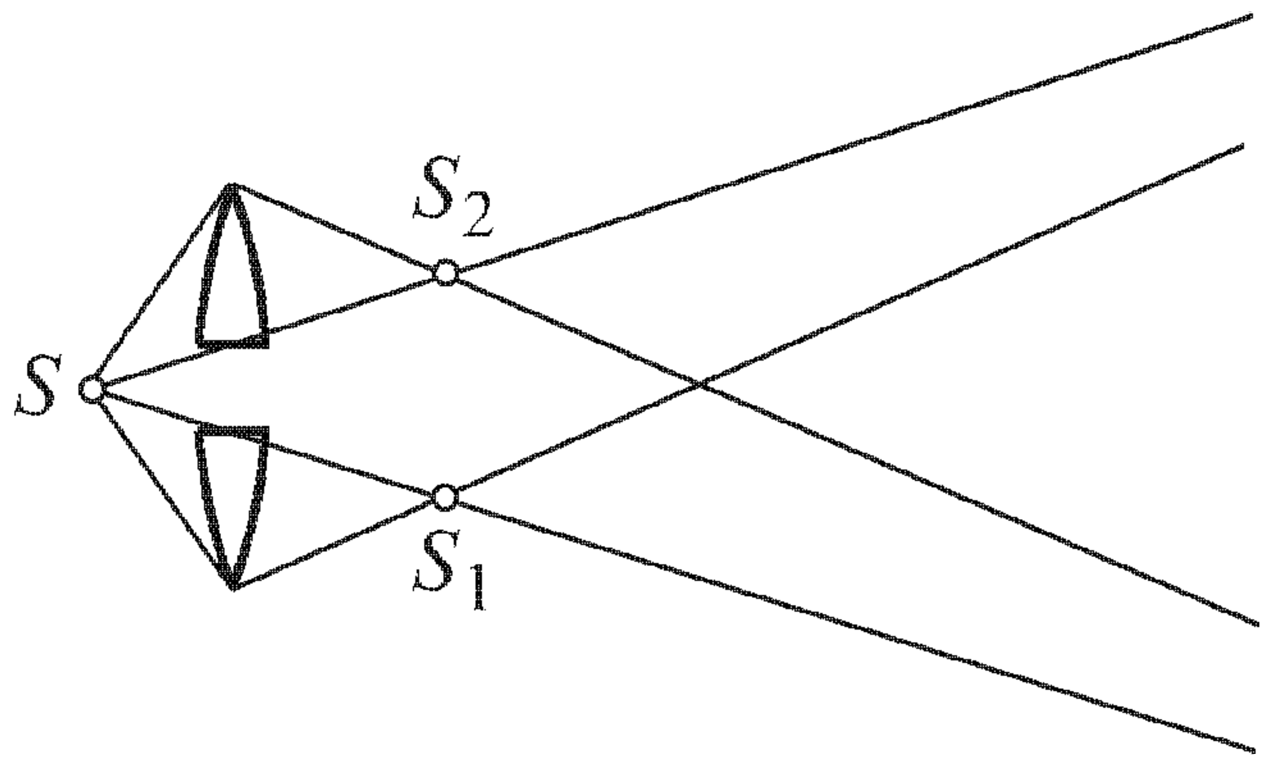


Рис. 118

4. Билинза Бийе. Собирательная линза разрезается пополам, и обе половинки ее раздвигаются. Образовавшаяся система называется *билинзой*. На билинзу направляется свет от щели S, параллельной плоскости разреза (рис. 118). В S_1 и S_2 получаются действительные изображения щели S. Пучки света, проходящие через S_1 и S_2, перекрываются. В местах перекрытия наблюдается интерференция. Если щель S поместить между билинзой и ее фокальной плоскостью, то изображения S_1 и S_2 будут мнимыми. В этом случае интерференция невозможна, так как световые пучки не перекрываются. Чтобы получить интерференцию и при таком расположении щели S, надо из центральной части линзы вырезать плоскопараллельный кусок стекла и обе половинки линзы сблизить между собой. Тогда пучки будут перекрываться, и интерференция станет возможной.

5. Зеркало Ллойда (1800–1881). Свет, исходящий из узкой ярко освещенной щели S_1 (рис. 119), отражается от полированной плоской поверхности черного стекла. На участке AB экрана, где перекрываются прямой и отраженный свет, наблюдаются полосы интерференции.

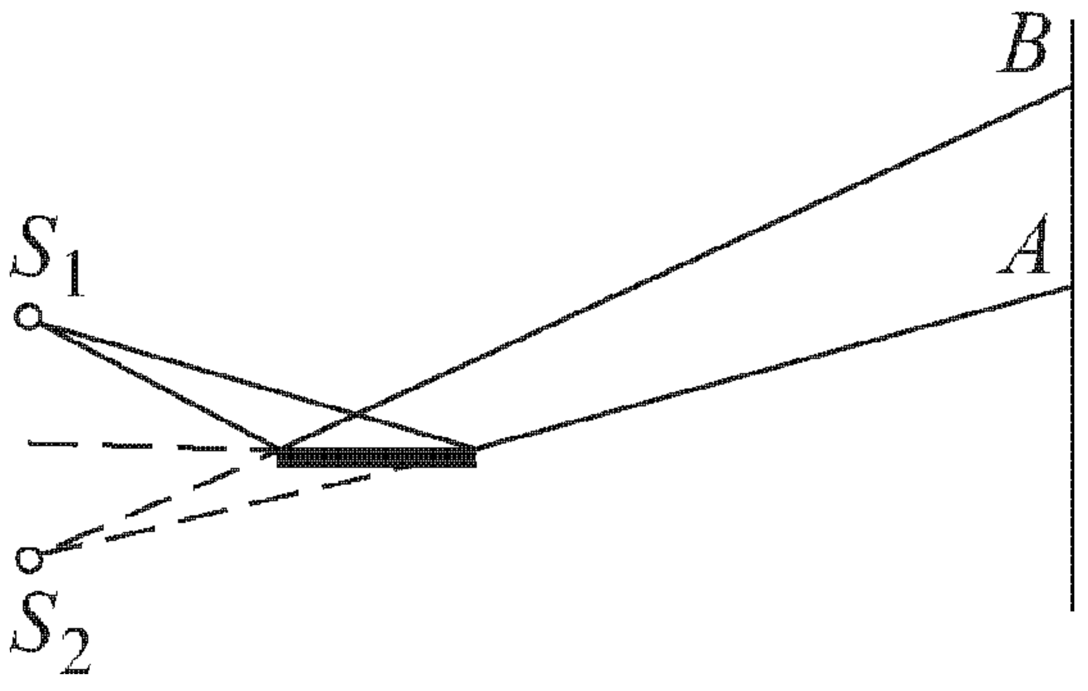


Рис. 119

6. Во всех описанных опытах полосы интерференции можно наблюдать на белом матовом экране. Можно пользоваться также матовым стеклом, наблюдая полосы интерференции на его задней стороне через лупу или микроскоп малого увеличения. Такой *объективный* способ наблюдения применял и Френель в первых опытах. Но он заметил, что полосы интерференции не исчезали, а становились более яркими, когда матовое стекло убиралось. Поэтому в дальнейших исследованиях Френель перешел к *субъективному* методу наблюдения интерференционных (и дифракционных) полос. Идея, на которой основан субъективный способ наблюдения полос, весьма проста. Допустим, что интерференционные полосы получены на задней плоскости П матового стекла. Интенсивность света в какой-либо точке P плоскости П определяется разностью фаз интерферирующих лучей, приходящих в эту точку. Отобразим теперь с помощью линзы плоскость П на сопряженную плоскость П$'$. Лучи, вышедшие из точки P, снова соберутся теперь в точке P'. Но так как оптические длины всех лучей между сопряженными точками одинаковы, то интерферирующие лучи придут в точку P' *с той же разностью фаз*, какой они обладали в точке P. Поэтому в плоскости П$'$ получится изображение не только плоскости П, но и всей системы интерференционных полос на ней. Линза как бы переносит интерференционную картину с плоскости П на плоскость П$'$. Так же действует и оптическая система глаза, с помощью которой интерференционные полосы получаются на сетчатке. Если пользоваться зрительной трубой, установленной на бесконечность, то таким путем будет наблюдаться интерференционная картина, получающаяся на бесконечном (т.е. очень далеком) расстоянии от глаза.

7. Интерференционный опыт Поля (1884–1976). Описанные выше установки отличаются малой светосилой, а потому мало пригодны для демонстрации интерференции света в большой аудитории. Опыт Поля свободен от этого недостатка. Здесь источник света S (рис. 120), которым может служить небольшая, но достаточно яркая ртутная лампа, помещается перед пластинкой слюды, толщина которой порядка нескольких сотых долей миллиметра. Свет отражается от передней и задней поверхностей пластинки с образованием двух мнимых изображений S_1 и S_2 источника S. Прямой свет от источника S загораживается ширмой. Получаются два широких когерентных пучка света с вершинами в точках S_1 и S_2. Падая на экран, потолок или

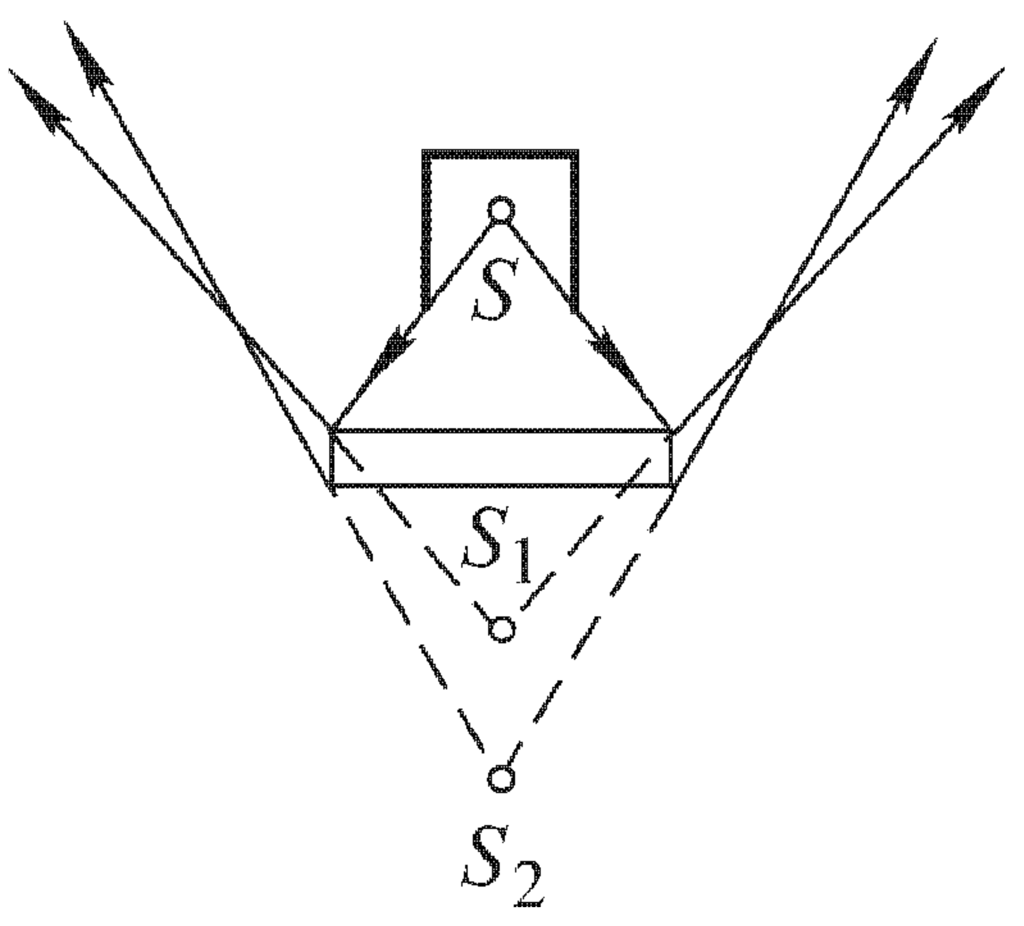


Рис. 120

стены аудитории, они дают при интерференции систему светлых и темных колец (см. п. 6 предыдущего параграфа). Благодаря большому количеству света кольца хорошо видны во всей аудитории. Преимущества метода Поля будут выяснены в § 28.

8. Опыт Меслина. Этот опыт является видоизменением опыта Бийе. Только в нем половинки разрезанной линзы не раздвигаются перпендикулярно к оптической оси, а смещаются вдоль нее, и притом на значительное расстояние (рис. 121). Половинки билинзы дают действительные изображения точечного источника S в точках S_1 и S_2 на главной оптической оси. Интерференционные полосы наблюдаются в заштрихованной области, где накладываются соответствующие пучки света, т.е. между изображениями S_1 и S_2. В плоскости, перпендикулярной к главной оптической оси, интерференционные полосы имеют вид концентрических полуколец с общим центром на этой оси.

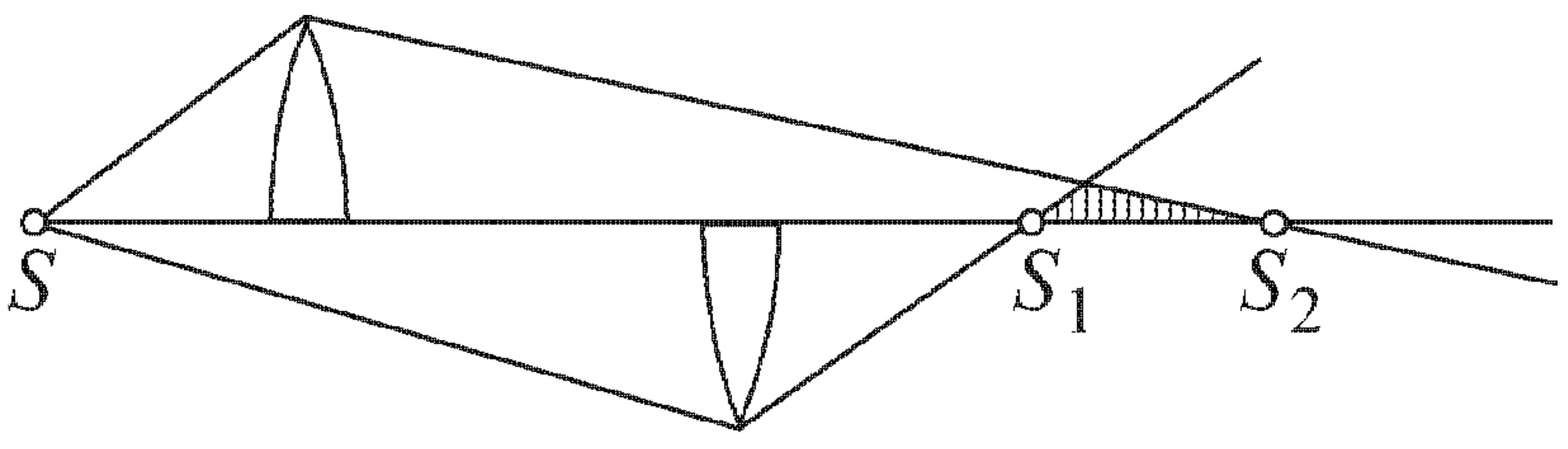


Рис. 121

Интересно, что центр колец *темный*. Явление связано с тем, что один пучок, прежде чем интерферировать со вторым, проходит через свой фокус S_1, тогда как другой пучок проходит через свой фокус S_2 позже. На этом основании можно прийти к заключению, что прохождение волны через фокус сопровождается изменением фазы на π.

Это легко понять на примере скалярной сферической волны, сходящейся к своему центру O и заполняющей полный телесный угол 4π. Представим поле этой волны в виде $E_1 = \frac{1}{r} f_1\left(t + \frac{r}{v}\right)$, где r — расстояние от центра O, а v — скорость распространения волны. Когда волна проходит через центр O, начинает образовываться расходящаяся сферическая волна $E_2 = \frac{1}{r} f_2\left(t - \frac{r}{v}\right)$. Волновое поле в пространстве теперь представится в виде

$$E = \frac{1}{r} f_1\left(t + \frac{r}{v}\right) + \frac{1}{r} f_2\left(t - \frac{r}{v}\right).$$

Но напряженность поля E нигде не может обращаться в бесконечность, в частности и при $r = 0$. Поэтому необходимо, чтобы в любой момент времени выполнялось соотношение $f_2(t) = -f_1(t)$, т.е. функция f_2 может отличаться от функции f_1 только знаком. Если колебания f_1 и f_2 синусоидальны, то в точке O их фазы будут противоположны. А это и значит, что при прохождении через фокус фаза волны меняется на π.

Из приведенного рассуждения следует, что на передовом фронте функция f_1 должна оставаться непрерывной. Если бы здесь она испытывала разрыв, то при подходе передового фронта к центру O напряженность поля E возрастала бы неограниченно. То же самое относится и к заднему фронту волны, а также к функции f_2.

ЗАДАЧИ

1. Как надо поместить собирательную линзу в опыте с бипризмой (см. рис. 117), чтобы ширина интерференционной полосы Δx не зависела от положения экрана? Найти Δx. Найти максимальное число N полос, которое может наблюдаться в этой установке, если оно получается при удалении экрана от бипризмы на расстояние L.

Ответ. $\Delta x = \dfrac{\lambda}{2(n-1)\beta}$; $N = \dfrac{4L(n-1)^2\beta^2}{\lambda}$.

2. При каком расстоянии L экрана от бипризмы в предыдущей задаче число интерференционных полос N будет максимальным, если расстояние между вершинами бипризмы равно l? Найти N. При каком положении экрана полосы пропадут?

Ответ. $N = l(n-1)\beta/\lambda$; $L = \dfrac{l}{4(n-1)\beta}$. Если экран удалить от бипризмы на расстояние L или больше, то полосы пропадут.

3. Найти число полос интерференции N в установке с бипризмой (см. рис. 117).

Ответ. $N = \dfrac{4ab}{a+b}\dfrac{(n-1)^2\beta^2}{\lambda}$.

4. В опыте с бипризмой Френеля экран для наблюдения интерференционных полос расположен перпендикулярно к оси установки. Нулевая полоса получается в центре экрана в вертикальной плоскости, проходящей через ребро бипризмы и ось установки. В какую сторону и на какую величину x сместится нулевая полоса, если освещающую щель немного сместить в сторону на величину h перпендикулярно к оси установки?

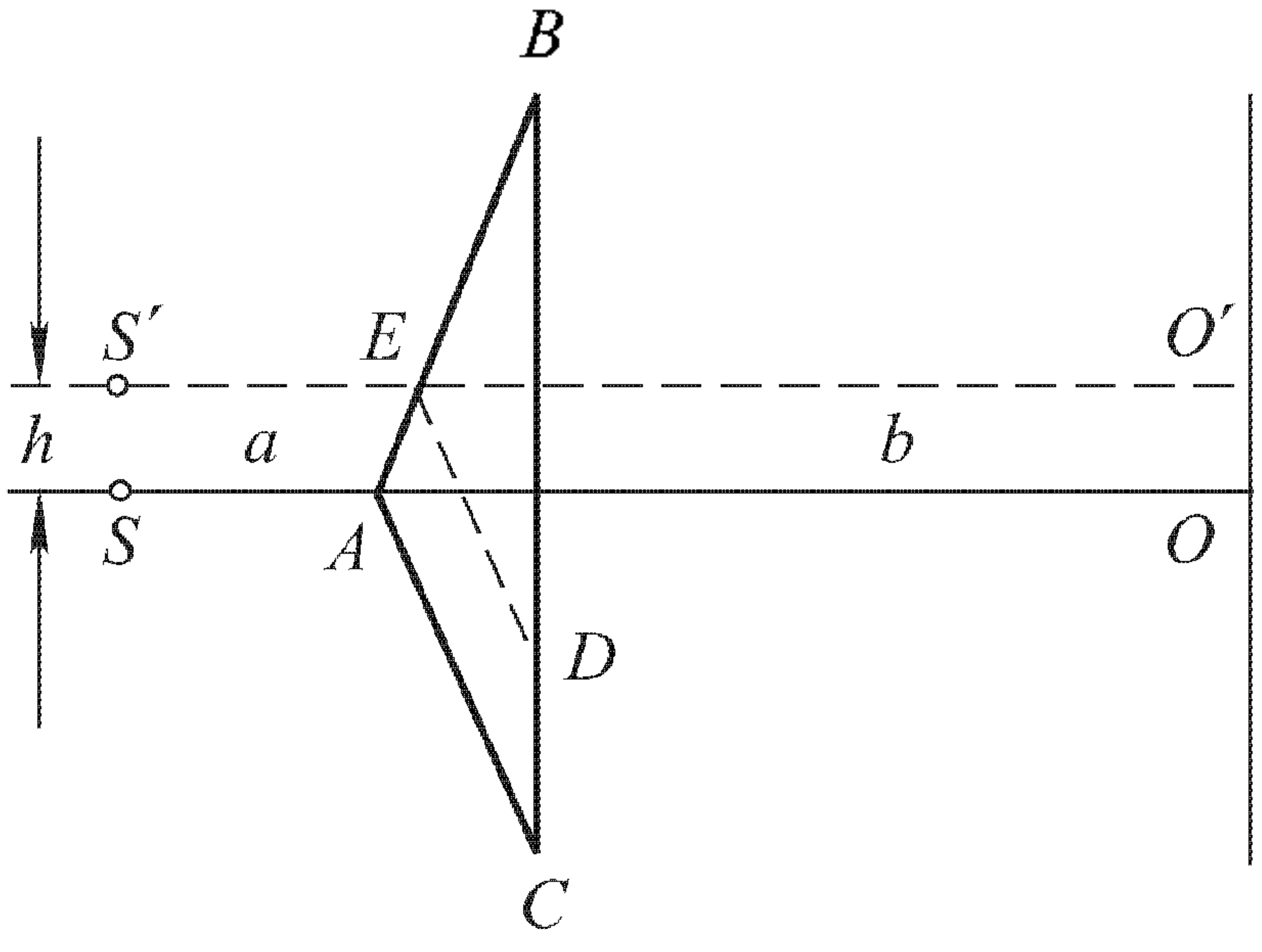


Рис. 122

Решение. При смещении источника из положения S в положение S' (рис. 122) действие бипризмы CAB сведется к действию ее части DEB и плоскопараллельной пластинки $CAED$. Суммируя смещения, вызываемые этими частями в отдельности, найдем, что центр интерференционной картины из прежнего положения O сместится вверх на расстояние $x = hb/a$.

§ 28. Влияние размеров источника света. Пространственная когерентность

1. Увеличение размеров источника света приводит к ухудшению *контрастности* интерференционных полос и даже к их полному исчезновению. Допустим сначала, что источник состоит из двух одинаковых некогерентных светящихся точек A и B, находящихся на малом расстоянии l друг от друга (рис. 123). Как было выяснено в § 26 (п. 10), для получения интерференции от одного источника A надо расщепить свет, исходящий от этого источника, на два пучка, попадающих на экран *различными путями*. То же самое надо сделать со светом от источника B. Конкретный способ расщепления пучка лучей на два в последующих рассуждениях не имеет никакого значения. На рис. 123 один пучок света попадает к месту интерференции P непосредственно от источников A и B, другой — после отражения от зеркала M_2N_2. Последующие рассуждения применимы и к криволинейным лучам. Однако, во избежание ненужных усложнений в написании формул, показатели преломления сред в точках A и B, а также в месте нахождения экрана, на котором наблюдается интерференция, предполагаются равными единице.

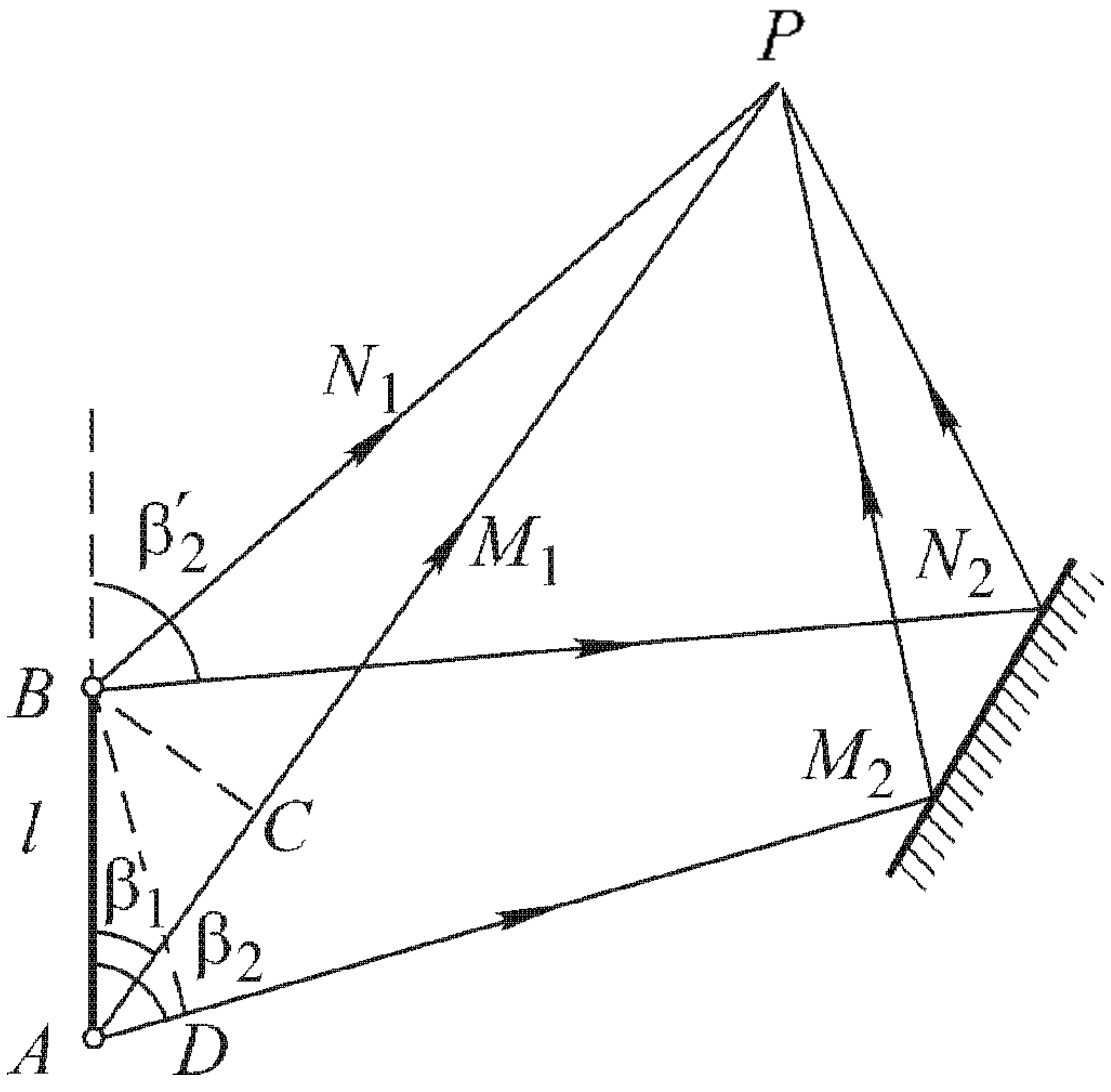


Рис. 123

Пусть AM_1P и AM_2P — два луча, исходящие из точки A под углами β_1 и β_2 к линии AB и сходящиеся в точке P, где они и интерферируют. Аналогичные лучи, исходящие из точки B, пусть будут BN_1P и BN_2P. Если расстояние $l = AB$ достаточно мало, то для разности оптических длин лучей AM_1P и BN_1P можно написать: $(AM_1P) - (BN_1P) = (AC) = l\cos\beta_1$. Аналогично для AM_2P и BN_2P: $(AM_2P) - (BN_2P) = (AD) = l\cos\beta_2$. Следовательно,

$$[(AM_1P) - (AM_2P)] - [(BN_1P) - (BN_2P)] = l(\cos\beta_1 - \cos\beta_2).$$

В первой квадратной скобке стоит оптическая разность хода лучей, исходящих из точки A. Она определяет характер интерференции этих лучей в точке P: максимум, минимум или какой-либо промежуточный случай. Аналогично, во второй квадратной скобке стоит оптическая разность хода лучей, исходящих из точки B, от которой зависит характер интерференции этих лучей. Разность этих величин

$$\Delta = l|\cos\beta_1 - \cos\beta_2|$$

характеризует фазовый сдвиг интерференционной картины от точечного источника A относительно интерференционной картины от точечного источника B. Она определяет результат наложения одной из этих картин на другую. Если величина Δ равна нулю или мала по сравнению с длиной волны λ, то максимумы одной картины накладываются на максимумы другой, а минимумы — на минимумы. Тогда происходит усиление интерференционных картин, а полосы интерференции наиболее контрастны: наблюдается плавный переход от яркой освещенности в максимумах к полной темноте в минимумах. При возрастании Δ контрастность полос сначала ухудшается. Когда $\Delta = \lambda/2$, т.е.

$$l|\cos\beta_1 - \cos\beta_2| = \lambda/2, \tag{28.1}$$

максимумы одной картины накладываются на минимумы другой, и полосы интерференции исчезают. При дальнейшем возрастании Δ полосы интерференции появляются вновь, причем интерференционная картина периодически повторяется от наибольшей контрастности до их полного исчезновения. При

$$\Delta \equiv l|\cos\beta_1 - \cos\beta_2| = m\lambda \tag{28.2}$$

(m — целое число) полосы столь же контрастны, что и при $\Delta = 0$. При

$$\Delta \equiv l|\cos\beta_1 - \cos\beta_2| = \left(m + \frac{1}{2}\right)\lambda \tag{28.3}$$

полосы пропадают, а экран освещается равномерно. Если Δ не превышает примерно $\lambda/4$, то контрастность полос лишь немного уступает контрастности при $\Delta = 0$. То же самое будет при $\Delta = (m+\alpha)\lambda$, где α — правильная дробь, не превышающая примерно $1/4$. Это несколько произвольное условие в дальнейшем мы будем рассматривать как *условие хорошей контрастности интерференционных полос.* Заметим еще, что в пределах принятой точности расчета угол β_2 можно отождествить с углом β_2', что и делается в дальнейшем без особых оговорок.

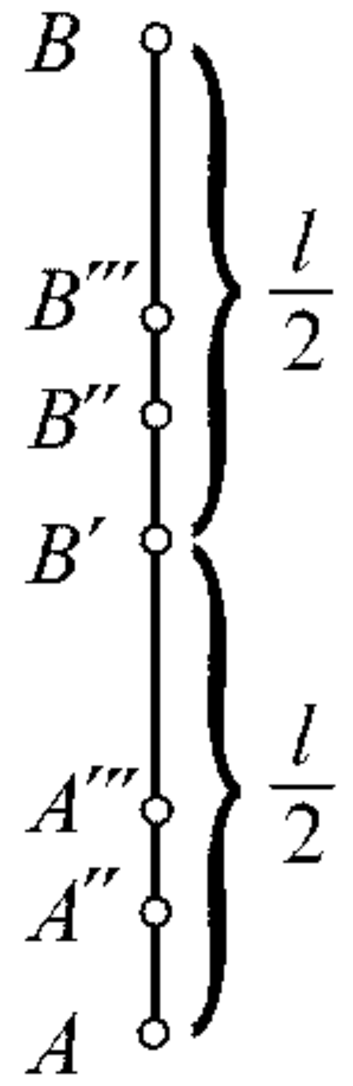


Рис. 124

2. Допустим теперь, что источник света протяженный и имеет форму равномерно светящейся линии AB длины l, все точки которой излучают свет некогерентно (рис. 124). Разобьем ее мысленно на бесконечное множество пар некогерентных точечных источников (A, B'), (A'', B''), ..., находящихся на расстоянии $l/2$ друг от друга. Применим к таким парам точечных источников результаты, полученные выше. Для этого надо расстояние l заменить на $l/2$. Если $\frac{\Delta}{2} = \frac{l}{2}|\cos\beta_1 - \cos\beta_2| = \frac{\lambda}{2}$, то интерференционных полос от каждой пары точечных источников не получится, они создадут на экране только светлый фон. То же самое будет, если размеры источника $l = AB$ увеличить в 2, 3, ... раз. Итак, если выполнено условие

$$\Delta \equiv l|\cos\beta_1 - \cos\beta_2| = m\lambda, \tag{28.4}$$

то протяженный источник создает на экране только освещенный фон без интерференционных полос (в противоположность тому, что было в случае пары точечных источников, когда при том же условии получалась наилучшая контрастность полос). Если $\Delta = (m + \alpha)\lambda$, где α — правильная дробь, то протяженный источник можно мысленно разбить на две части, длины которых относятся как $m : \alpha$. Меньшая часть источника даст интерференционные полосы на светлом фоне, создаваемом большей частью того же источника. От этого контрастность полос уменьшается. Чем больше m, тем хуже контрастность. При больших m интерференционные полосы практически наблюдаться не будут. Условие хорошей контрастности интерференционных полос в случае протяженного источника света можно ориентировочно представить в виде

$$l|\cos\beta_1 - \cos\beta_2| \leqslant \lambda/2. \tag{28.5}$$

Если крайние интерферирующие лучи выходят из какой-либо точки протяженного источника симметрично по отношению к перпендикуляру к линии AB, проведенному из этой точки, т.е. $\beta_2 = \pi - \beta_1$, то $\cos\beta_2 = -\cos\beta_1$, и предыдущее условие можно представить в виде

$$l\sin(\Omega/2) \leqslant \lambda/4, \tag{28.6}$$

где Ω — угол между этими крайними лучами. Его часто называют *углом интерференции*. Следует, однако, иметь в виду, что условие (28.6) можно применять только при наличии той *симметрии*, которая была использована при его выводе. Задания только одного угла интерференции (вопреки распространенному противоположному утверждению) совершенно недостаточно, чтобы в общем виде записать условие хорошей контрастности интерференционных полос. Для этого надо ввести два угла β_1 и β_2, как это и было сделано выше.

Результаты, полученные выше для точечных источников, без всяких затруднений распространяются на источники света, имеющие форму коротких параллельных прямолинейных отрезков, смещенных один относительно другого в поперечном направлении на расстояние l и параллельных плоскости экрана, в средней части которого наблюдается интерференция. (Под средней частью здесь понимается малая окрестность экрана вблизи линии, вдоль которой экран пересекается с плоскостью, проведенной через середину источников перпендикулярно к их длине.)

При том же расположении справедливы и результаты (28.4) и (28.5) для некогерентных протяженных источников, имеющих форму прямоугольных полосок, если только под l понимать теперь их ширину. Такими источниками в оптике обычно служат прямоугольные щели, ярко освещаемые широкими пучками лучей. Если выполнено условие

$$l|\cos\beta_1 - \cos\beta_2| \ll \lambda,$$

то щель действует как бесконечно тонкий линейный источник света.

3. Таким образом, для получения интерференционных полос от двух источников света недостаточно, чтобы эти источники состояли из попарно когерентных точечных источников. (Примером могут служить источник и его оптическое изображение или два оптических изображения одного и того же источника.) Даже в случае строго монохроматического света необходимо еще, чтобы размеры источников не превосходили определенного предела, зависящего от взаимного расположения и расстояния между ними, а также от положения экрана, предназначенного для наблюдения интерференционных полос.

Два источника, размеры и взаимное расположение которых позволяют наблюдать интерференционные полосы (при необходимой степени монохроматичности света), называются *пространственно когерентными.* Если же интерференционных полос (даже при идеальной монохроматичности света) получить нельзя, то источники называются *пространственно некогерентными* (см. далее п. 8).

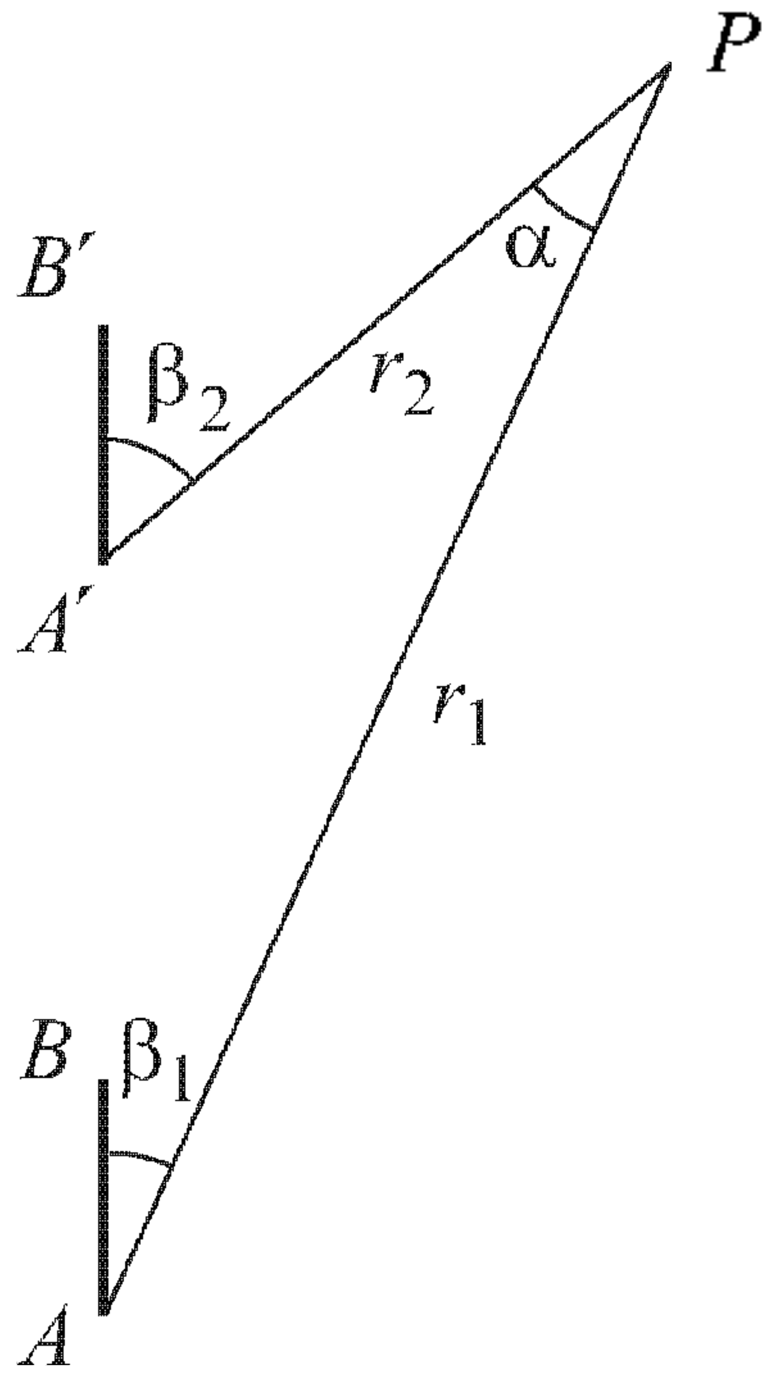


Рис. 125

4. Посмотрим теперь, при каких условиях будет соблюдаться пространственная когерентность в классических интерференционных опытах, описанных в предыдущем параграфе. Во всех опытах источники света линейные (за исключением опыта Меслина, где они точечные, так что вопрос о пространственной когерентности здесь практически не возникает). При этом в опытах с зеркалами и бипризмой Френеля, а также в установках Бийе и Поля взаимное расположение когерентных источников AB и $A'B'$ «прямое», т.е. соответствует рис. 125, а в опыте с зеркалом Ллойда — «обратное», как представлено на рис. 126 (стр. 221). На этих рисунках соответствующие когерентные пары точечных источников обозначены через (A, A'), (B, B') и т.д. При этом в установках Френеля, Ллойда и Бийе, где источниками служат освещаемые щели, последние предполагаются перпендикулярными к плоскости рисунка, а $l = AB$ означает ширину щели. В установке Поля линейные источники AB и $A'B'$ (например, изображение ртутной лампы, имеющей форму небольшого цилиндра) ориентированы вертикально, так что здесь AB означает длину l источника света.

5. Начнем с прямого расположения (см. рис. 125). В этом случае $\beta_2 - \beta_1 = \alpha$. Ввиду малости угла α схождения интерферирующих лучей можем написать

$$\cos\beta_1 - \cos\beta_2 = \sin\beta\Delta\beta = \alpha\sin\beta,$$

где с принятой точностью расчета под β можно понимать любой из углов β_1 и β_2 или любое промежуточное значение между ними. Условие (28.5) хорошей контрастности интерференционных полос принимает вид

$$l \leqslant \lambda/(2\alpha\sin\beta), \tag{28.7}$$

или, ввиду (26.13),

$$l \leqslant \Delta x/(2\sin\beta). \tag{28.8}$$

В установках Френеля и Бийе углы β_1 и β_2 практически не отличаются от $\pi/2$, так что в этом случае $l \leqslant \Delta x/2$, т.е. ширина щели не должна превосходить примерно половину ширины интерференционной полосы (предполагается, что плоскость экрана перпендикулярна к оси установки CO, см. рис. 116).

В интерференционном опыте Поля (см. рис. 120), напротив, углы β составляют 20–50°, а интерференционные полосы, если они наблюдаются на потолке аудитории, имеют форму колец. Для упрощения вычислений заменим пластинку слюды двумя отражающими параллельными плоскостями, расстояние между которыми равно толщине h этой пластинки. Тогда расстояние между соответствующими точками источников (например, между A и A' или B и B') будет $2h$, так что $2h\sin\beta = r\alpha$, где r в нашем приближении равно r_1 или r_2 либо любой промежуточной величине между r_1 и r_2. Подставляя отсюда значение α в (28.7), получаем условие пространственной когерентности:

$$l \leqslant r\lambda/(4h\sin^2\beta). \tag{28.9}$$

В опыте Поля толщина пластинки слюды h должна быть очень мала, так что мнимые источники AB и $A'B'$ сдвинуты один относительно другого на ничтожную величину $2h$, во много раз меньшую длины l самого источника. Допустим, например, что $\lambda = 500$ нм $= 5\cdot10^{-5}$ см, $h = 0{,}05$ мм, $r = 8$ м, $\beta = 30^\circ$. Тогда, согласно формуле (28.9), должно быть $l \leqslant 8$ см. Таким образом, для получения хорошей контрастности полос источник света может иметь сравнительно большие размеры, а потому его можно взять светосильным. В этом основное преимущество установки Поля при демонстрации явлений интерференции. Другое преимущество — большие апертуры интерферирующих пучков, позволяющие получать полосы интерференции на большой площади.

6. Перейдем теперь к обратному расположению когерентных источников AB и $A'B'$ (рис. 126), реализующемуся в опыте Ллойда. В этом случае в середине экрана O условие (28.5) выполняется при любой ширине щели l. При другом расположении точки наблюдения P

$$\cos\beta_1 = (AC + OP)/r_1, \quad \cos\beta_2 = -(OP - A'C)/r_2.$$

Так как $AC = A'C$, то, пренебрегая в знаменателях различием между r_1 и r_2, отсюда получим: $\cos\beta_1 - \cos\beta_2 = 2x/r$, где $x = OP$ — координата точки P. Поскольку интерферирующие лучи почти перпендикулярны к плоскости экрана, условие хорошей

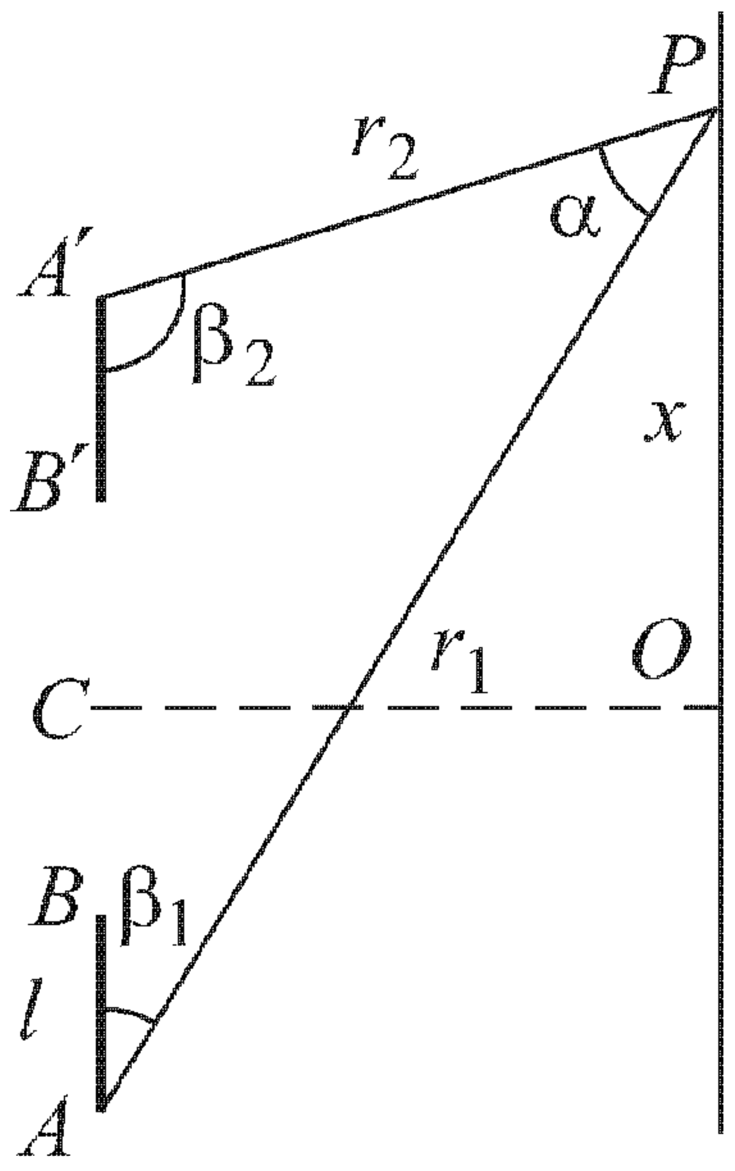


Рис. 126

контрастности интерференционных полос запишется в виде $2lx/r \leqslant \lambda/2$, т.е.

$$x \leqslant r\lambda/(4l). \tag{28.10}$$

На вдвое большем расстоянии интерференционные полосы пропадут. Полное число N полос, которые будут наблюдаться до такого расстояния, найдется делением $2x$ на ширину полосы $\Delta x = \lambda/\alpha$. Это дает

$$N = \frac{r\alpha}{2l} = \frac{d}{l}, \tag{28.11}$$

где $2d = r\alpha$ — расстояние между источниками света. Это выражение дает оценку наивысшего порядка интерференции в монохроматическом свете в установках типа опыта Ллойда.

7. Рассмотрим, наконец, интерференционный опыт Юнга (рис. 127). Источником света служит прямоугольная полоска ширины l (например, ярко освещенная щель). Из каждой точки этой полоски, например, точки A, лучи идут к щелям S_1 и S_2 экрана под углами β_1 и β_2. Если опять пренебречь в знаменателях различием между r_1 и r_2, то можно написать: $\cos\beta_1 - \cos\beta_2 = d/r$, где d — расстояние между серединами щелей S_1 и S_2. Тогда условие (28.5) переходит в $ld/r \leqslant \lambda/2$, или

$$\varphi \leqslant \lambda/(2d), \tag{28.12}$$

где $\varphi = l/r$ — угловой размер источника AB, если его рассматривать с места расположения щелей S_1 и S_2 (так как в опыте Юнга углы β_1 и β_2 почти не отличались от $\pi/2$). Когда $\varphi = \lambda/d$, интерференционные полосы пропадают.

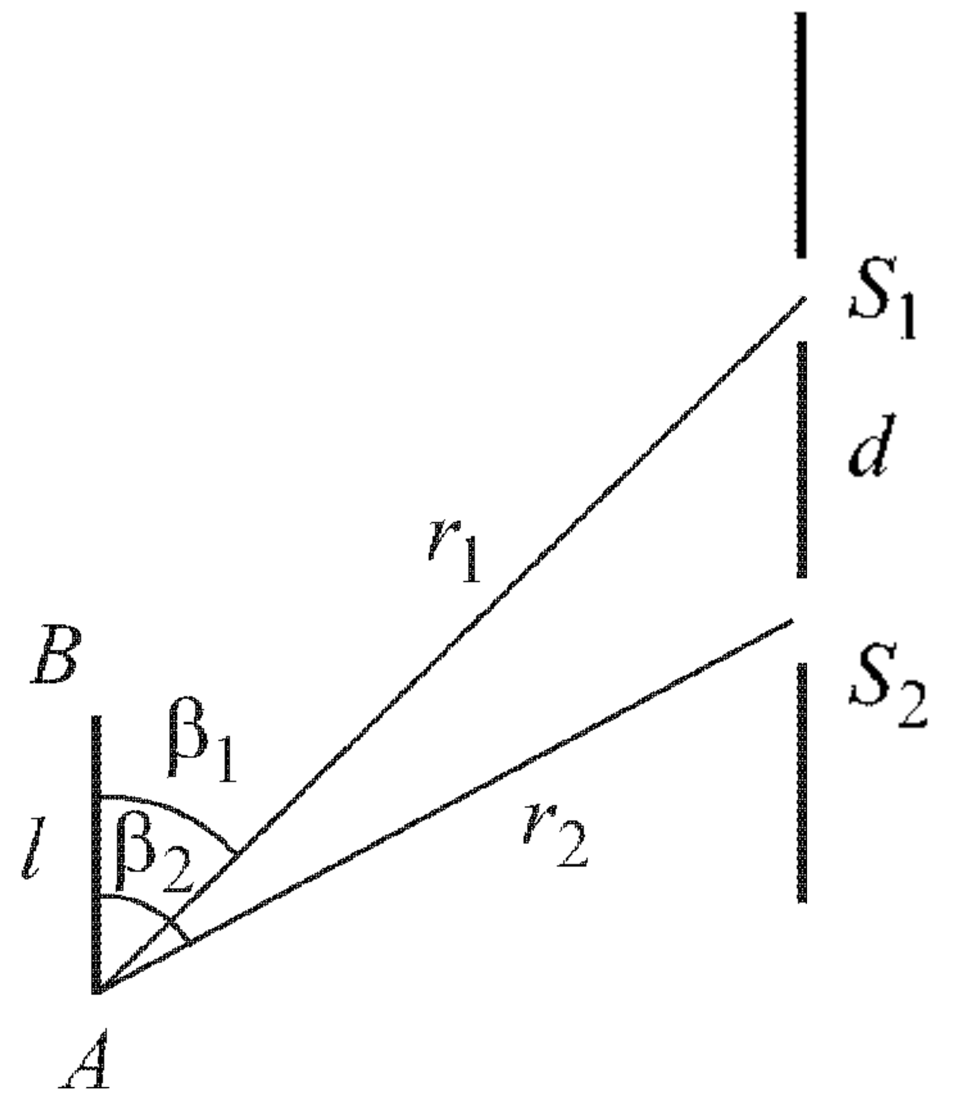


Рис. 127

Допустим, что свет направляется к щелям S_1 и S_2 непосредственно от Солнца, как это было в опыте Гримальди (см. § 27, п. 1; в опыте Гримальди источниками S_1 и S_2 служили два рядом расположенных небольших отверстия). Угловой размер Солнца $\varphi \approx 30' = 0{,}0087$ рад, длину волны примем равной $\lambda = 550$ нм. Тогда для получения каких бы то ни было интерференционных полос необходимо выполнение условия $d < \lambda/\varphi \approx \approx 6 \cdot 10^4$ нм $= 0{,}06$ мм. Такое условие в опытах Гримальди, конечно, не могло соблюдаться, а потому не могла наблюдаться и интерференция света.

8. Более важным является следующее обобщение. Допустим, что свет, освещающий отверстия S_1 и S_2, монохроматичен. Будет или не будет в этих условиях наблюдаться интерференция пучков света, прошедших через малые отверстия S_1 и S_2, при заданном расстоянии d между ними, зависит от степени углового расхождения лучей, освещающих эти отверстия. Пусть, например, источник имеет форму диска, плоскость которого параллельна плоскости экрана, на котором сделаны отверстия S_1 и S_2. Если φ — угловой диаметр диска, то интерференция будет наблюдаться, когда отверстия S_1 и S_2 можно покрыть кругом с диаметром $d < \lambda/\varphi$. Если же этого сделать нельзя, то интерференции не будет. В первом случае говорят, что световые пучки, прошедшие через отверстия S_1 и S_2, *пространственно когерентны*, а во втором — *пространственно некогерентны*. Минимальная площадь поперечного сечения пучка

$$\sigma = \pi\lambda^2/(4\varphi^2), \tag{28.13}$$

обладающая этим свойством, называется *площадью пространственной когерентности* освещающего света. Когда свет строго монохроматичен, то все нарушения когерентности носят чисто пространственный характер, т.е. связаны с различием направлений световых лучей. Для строго плоских волн все направления лучей одинаковы $\varphi = 0$), так что площадь когерентности σ становится бесконечной. По мере удаления от источника угловое расхождение лучей, попадающих в прибор, уменьшается, а их пространственная когерентность повышается. Примером могут служить звезды. Несмотря на громадные линейные размеры звезд, свет от них доходит до нас с высокой степенью пространственной когерентности. Световые лучи от лазеров характеризуются высокой направленностью, недоступной никаким другим земным источникам света. Именно эту особенность лазерного излучения имеют в виду, когда говорят о его высокой пространственной когерентности.

ЗАДАЧИ

1. Для характеристики контрастности интерференционных полос Майкельсон ввел функцию

$$V = \frac{I_{\text{макс}} - I_{\text{мин}}}{I_{\text{макс}} + I_{\text{мин}}}, \tag{28.14}$$

называемую *видностью* (или *видимостью*) их. Определить видность V в случае интерференции двух одинаково интенсивных монохроматических пучков света, попадающих на экран по различным путям от точечного источника света.

Решение. Согласно (26.7), интенсивность света на экране при освещении обоими источниками дается выражением $I = 2I_1(1 + \cos\delta)$, где I_1 — интенсивность, создаваемая одним из них. При смещении точки наблюдения по экрану из-за изменения разности фаз δ между интерферирующими пучками интенсивность I меняется от $I_{\text{макс}} = 4I_1$ до $I_{\text{мин}} = 0$, так что $V = 1$.

2. Решить ту же задачу для источника света, состоящего из двух одинаковых некогерентных светящихся точек A и B, расположенных на расстоянии l друг от друга (см. рис. 123). Интерференционные полосы наблюдаются на удаленном экране.

Решение. Интенсивности света в интерференционных картинах на экране, создаваемые в отдельности светящимися точками A и B, равны соответственно $I_A = 2I_1(1+\cos\delta_1)$, $I_B = 2I_1(1+\cos\delta_2)$, где δ_1, δ_2 и I имеют такой же смысл, что и в предыдущей задаче. Результирующая интенсивность

$$I = I_A + I_B = 4I_1\left(1 + \cos\frac{\delta_1 - \delta_2}{2}\cos\frac{\delta_1 + \delta_2}{2}\right).$$

Величина $\delta_1 - \delta_2$ есть разность фаз между обеими интерференционными картинами. Согласно вычислениям, приведенным в п. 1,

$$\delta \equiv |\delta_1 - \delta_2| = k\Delta = kl|\cos\beta_1 - \cos\beta_2|,$$

где $k = 2\pi/\lambda$ — волновое число. При перемещении по экрану точки наблюдения изменения фаз δ_1 и δ_2 происходят быстро и определяют ширину интерференционных полос. Напротив, разность фаз $\delta \equiv |\delta_1 - \delta_2|$ изменяется медленно, оставаясь практически постоянной на протяжении многих интерференционных полос. Она определяет видность, т.е. контрастность полос.

Максимальные и минимальные интенсивности будут

$$I_{\text{макс}} = 4I_1\left(1 + \left|\cos\frac{\delta}{2}\right|\right), \quad I_{\text{мин}} = 4I_1\left(1 - \left|\cos\frac{\delta}{2}\right|\right),$$

а потому

$$V = \left|\cos k\frac{\Delta}{2}\right| = \left|\cos\frac{\pi\Delta}{\lambda}\right|. \tag{28.15}$$

При изменении Δ, т.е. расстояния l между светящимися точками A и B, видность V периодически изменяется от 1 до 0 (рис. 128).

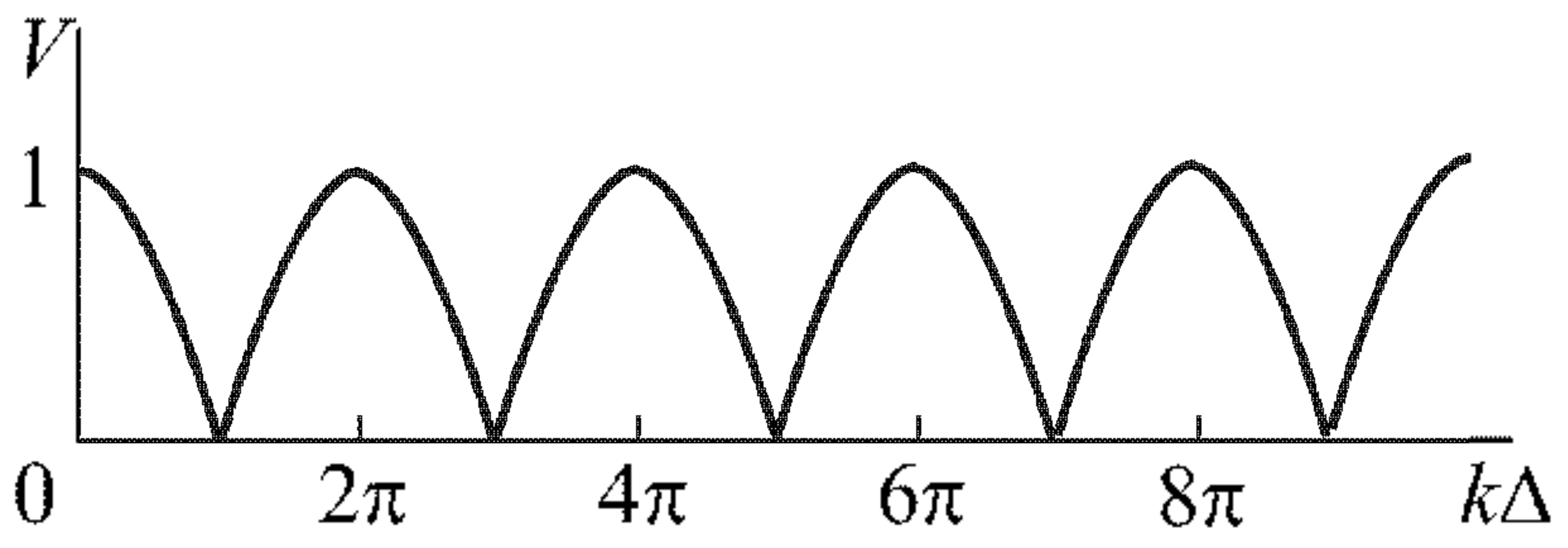


Рис. 128

3. Найти видность интерференционных полос, когда источником света является прямолинейный отрезок длины l, все точки которого излучают некогерентно (см. рис. 124).

Решение. Направим ось Y вдоль излучающего отрезка, поместив начало координат O в его середине. Пусть φ — разность фаз между двумя лучами, исходящими из точки O и попадающими в точку наблюдения P по различным путям. Разность фаз между такими же лучами, исходящими из другой точки отрезка с координатой y и приходящими в ту же точку P, будет

$$\delta = ky(\cos\beta_1 - \cos\beta_2) + \varphi = (k\Delta/l)y + \varphi.$$

Лучи, исходящие по различным путям из отрезка dy, при интерференции на экране создают интенсивность

$$dI = (1 + \cos\delta)\,dy$$

(в условных единицах). Так как точки источника излучают некогерентно, то полная интенсивность будет

$$I = \int\limits_{-l/2}^{+l/2}\left[1 + \cos\left(\frac{k\Delta}{l}y + \varphi\right)\right]dy = l + \frac{l}{k\Delta}\left[\sin\left(\frac{k\Delta}{2} + \varphi\right) + \sin\left(\frac{k\Delta}{2} - \varphi\right)\right].$$

При смещении вдоль экрана будет меняться фаза φ. Значения фазы φ, при которых интенсивность I экстремальна, найдутся из условия

$$\cos\left(\frac{k\Delta}{2} + \varphi\right) = \cos\left(\frac{k\Delta}{2} - \varphi\right),$$

откуда $\varphi = m\pi$, где m — целое число;

$$I = l + l\,\frac{\sin(k\Delta/2)}{k\Delta/2} \quad \text{при четном } m,$$

$$I = l - l\,\frac{\sin(k\Delta/2)}{k\Delta/2} \quad \text{при нечетном } m.$$

Следовательно,

$$I_{\text{макс}} = l + l\,\frac{\sin(k\Delta/2)}{k\Delta/2}, \quad I_{\text{мин}} = l - l\,\frac{\sin(k\Delta/2)}{k\Delta/2},$$

$$V = \frac{\sin(k\Delta/2)}{k\Delta/2}. \tag{28.16}$$

Кривая видности представлена на рис. 129. Из рисунка видно, как быстро с увеличением длины l источника уменьшаются максимумы на кривой видности. Той же кривой представляется видность интерференционных полос в опытах типа зеркал и бипризмы Френеля (если только точки щелевого источника света излучают некогерентно). В этом случае l означает ширину щели.

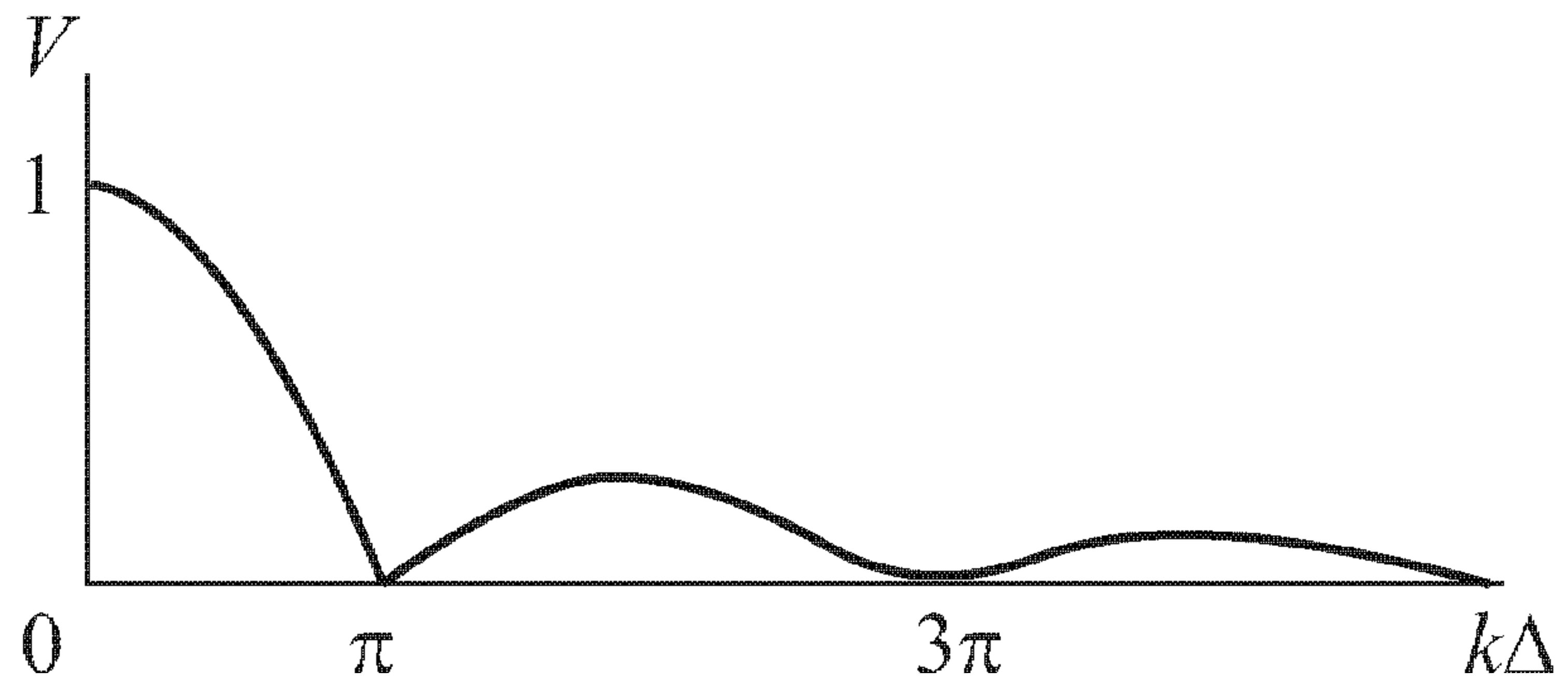


Рис. 129

4. Решить предыдущую задачу в предположении, что все точки источника излучают когерентно и притом в одинаковых фазах.

Ответ.

$$V = \frac{2\left|\dfrac{\sin(k\Delta/2)}{k\Delta/2}\right|}{1 + \left(\dfrac{\sin(k\Delta/2)}{k\Delta/2}\right)^2}. \tag{28.17}$$

§ 29. Спектральное разложение

1. До сих пор интерференция исследовалась только в идеальном случае монохроматического света. Интерференцию в немонохроматическом свете можно исследовать, разлагая свет по теореме Фурье на монохроматические составляющие. Если волновое поле в точке наблюдения описывается периодической функцией $E = E(t)$ с основным периодом τ и основной частотой $\Omega = 2\pi/\tau$, то его можно представить в виде вещественной части *ряда Фурье*:

$$E(t) = \sum_{n=0}^{\infty} a_n e^{in\Omega t}, \tag{29.1}$$

коэффициенты которого определяются выражением

$$a_n = \frac{1}{\tau} \int\limits_{-\tau/2}^{\tau/2} E(t) e^{-in\Omega t}\, dt \tag{29.2}$$

(см. т. III, § 128; конечно, в случае света постоянного слагаемого с коэффициентом a_0 не будет). Средняя за период объемная плотность энергии колебаний (в условных единицах) будет

$$\overline{w} = \frac{1}{\tau} \int\limits_{-\tau/2}^{\tau/2} |E|^2\, dt = \sum_{n=0}^{\infty} |a_n|^2 = \sum_{n=0}^{\infty} w_n. \tag{29.3}$$

Она равна сумме средних плотностей энергии монохроматических колебаний, из которых складывается результирующее колебание. То же заключение справедливо и для интенсивности колебаний, если понимать под интенсивностью усредненную по периоду τ любую энергетическую величину, характеризующую поле излучения в рассматриваемой точке пространства.

Полученный результат остается приближенно верным и для случая, когда функция $E(t)$ не периодична, а представляется суперпозицией монохроматических колебаний, частоты которых распределены по спектру совершенно произвольно. Только в этом случае усреднение надо производить не по периоду τ (которого теперь не существует), а по времени, весьма большому по сравнению с периодами всех монохроматических колебаний, входящих в суперпозицию. Результат приближенно верен и в случае суперпозиции почти гармонических колебаний с произвольными частотами, например для света, состоящего из узких спектральных линий.

2. Если функция $E(t)$ не периодична, то она представляется не рядом, а *интегралом Фурье*. Для возможности такого представления на функцию $E(t)$ приходится накладывать различные (достаточные) ограничения, например требовать, чтобы она была *абсолютно интегрируема* во всем бесконечном интервале $(-\infty, +\infty)$, т.е. чтобы сходился интеграл

$$\int\limits_{-\infty}^{+\infty} |E(t)|\, dt.$$

Это обстоятельство, однако, в физике не создает никаких существенных затруднений, даже в тех случаях, когда вводят функции, не обращающиеся в нуль на бесконечности, т.е. не удовлетворяющие требованию абсолютной интегрируемости.

Действительно, пусть $E(t)$ — такая функция. Разделим t на интервалы времени, достаточно длительные по сравнению с периодами световых колебаний. Световое поле на каждом из таких интервалов $(t_0, t_0+\tau)$ и его воздействие на приемник при любом значении t_0 практически совсем не зависят от полей на соседних интервалах. Поэтому при рассмотрении света только на интервале $(t_0, t_0+\tau)$ функцию $E(t)$ вне рассматриваемого интервала можно заменить любой другой функцией. В частности, ее можно периодически продолжить за пределы интервала $(t_0, t_0+\tau)$ с периодом τ. Но тогда для представления функции $E(t)$ в интервале $(t_0, t_0+\tau)$ можно воспользоваться рядом Фурье (29.1). При этом, ввиду малости частоты $\Omega = 2\pi/\tau$, целесообразно ввести обозначения $\Delta\omega = \Omega$, $\omega_n = n\Omega$. Тогда

$$E(t) = \sum a_n e^{i\omega_n t} = \sum \frac{a_n}{\Omega} e^{i\omega_n t} \Delta\omega.$$

После аппроксимации суммы интегралом получаем

$$E(t) = \int\limits_0^\infty a(\omega) e^{i\omega t}\, d\omega, \tag{29.4}$$

где $a(\omega) = a_n/\Omega$, или с учетом (29.2)

$$a(\omega) = \frac{1}{\tau\Omega} \int\limits_{-\tau/2}^{\tau/2} E(t) e^{-i\omega t}\, dt = \frac{1}{2\pi} \int\limits_{-\tau/2}^{\tau/2} E(t) e^{-i\omega t}\, dt. \tag{29.5}$$

Интеграл Фурье получается из этих формул, если в последнем выражении конечные пределы заменить на бесконечные $-\infty$ и $+\infty$ (см. т. III, § 128). Однако здесь мы не будем делать этого, оставляя время τ неопределенным. При физической постановке задач всегда можно достигнуть необходимой точности, выбирая τ достаточно большим. Таким путем достигается то преимущество, что формулу (29.4) в каждом интервале длительностью τ можно будет применять и для функций $E(t)$, не интегрируемых абсолютно, например к плоским волнам постоянной интенсивности, не ограниченным во времени. При этом выражения (29.4) на разных интервалах времени τ (если τ выбрать достаточно большим), вообще говоря, не будут когерентны.

В случае интеграла Фурье формула (29.3) заменится на

$$\int\limits_{-\tau/2}^{\tau/2} |E|^2\, dt = \tau \sum |a(\omega_n)\Omega|^2.$$

Если аппроксимировать сумму интегралом и учесть, что $\tau\Omega = 2\pi$, то получится

$$\int\limits_{-\tau/2}^{\tau/2} |E|^2\, dt = 2\pi \int\limits_0^\infty |a(\omega)|^2\, d\omega. \tag{29.6}$$

О физическом смысле этого соотношения говорится ниже в п. 5.

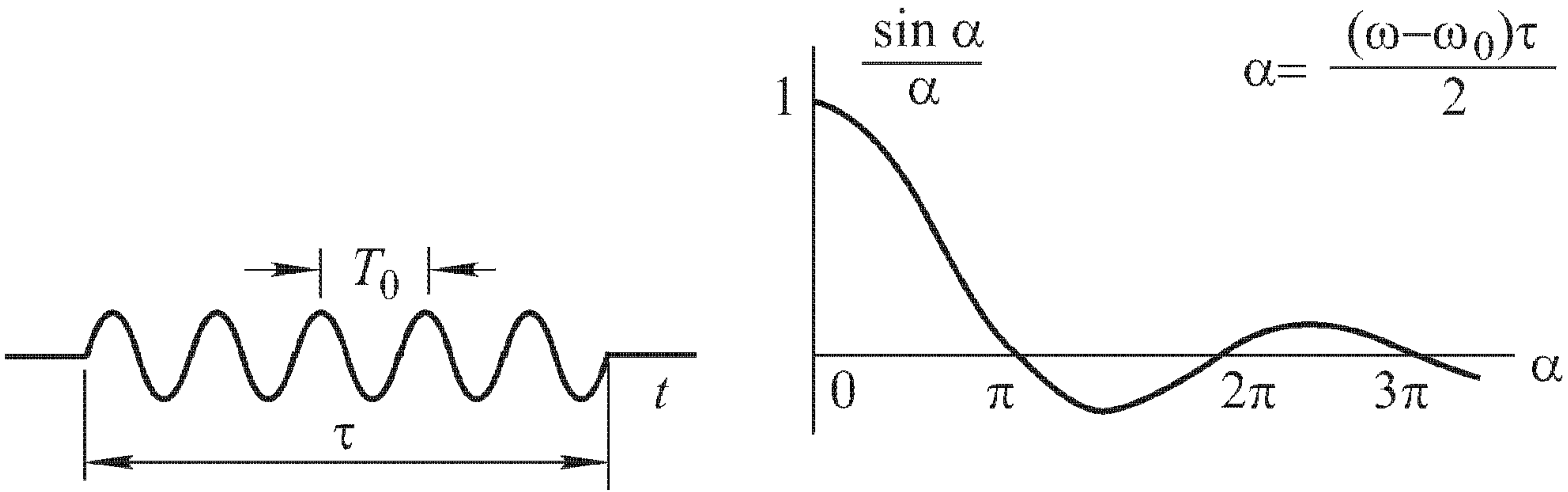
3. Приведем пример спектрального разложения, приводящий к важным обобщениям. Бесконечно длящееся синусоидальное колебание является идеализацией ограниченного ряда, или цуга синусоидальных волн, представленного на рис. 130 *а* «оборванной синусоидой». Пусть T_0 — «период», а ω_0 — «частота» этой «синусоиды». Тогда в комплексной форме

$$a(\omega) = \frac{1}{2\pi} \int\limits_{-\tau/2}^{\tau/2} \frac{1}{i} e^{i(\omega_0 - \omega)t}\, dt = \frac{2}{\pi i} \frac{\sin[(\omega - \omega_0)\tau/2]}{(\omega - \omega_0)\tau/2}.$$

В вещественной форме

$$E(t) = \frac{2}{\pi} \int\limits_0^\infty \frac{\sin[(\omega - \omega_0)\tau/2]}{(\omega - \omega_0)\tau/2} \sin\omega t\, d\omega. \tag{29.7}$$

График функции $\frac{\sin\alpha}{\alpha}$, где $\alpha = \frac{(\omega - \omega_0)\tau}{2}$, приведен на рис. 130 *б*. Таким образом, оборванный цуг волн, изображенный на рис. 130 *а*, может быть представлен суперпозицией бесконечного множества синусоид, частоты которых непрерывно заполняют бесконечный интервал $0 < \omega < +\infty$. Впрочем, основное значение имеет только интервал $-\frac{\pi}{2} < \alpha < +\frac{\pi}{2}$, или интервал $\omega_0 - \frac{\pi}{\tau} < \omega < \omega_0 + \frac{\pi}{\tau}$, где амплитуды колебаний велики. На всех остальных участках амплитуды малы. Если ими пренебречь, то можно сказать, что весь спектр частот практически сосредоточен в пределах интервала шириной $\Delta\omega$, который удовлетворяет условию



а *б*

Рис. 130

$$\Delta\omega \cdot \tau \gtrsim 2\pi. \tag{29.8}$$

Если ввести обычную частоту $\nu = \omega/(2\pi)$, то

$$\Delta\nu \cdot \tau \gtrsim 1. \tag{29.8a}$$

Это важное соотношение между шириной спектра $\Delta\omega$ (или $\Delta\nu$) и длительностью цуга τ имеет общий характер. Его можно также уяснить на следующем простом примере. Рассмотрим множество синусоид с одинаковыми амплитудами, но различными частотами, непрерывно и равномерно заполняющими интервал $\Delta\omega$. Пусть в точке $t = 0$ фазы всех синусоид совпадают, а следовательно, амплитуда колебаний максимальна. При $t = \tau/2$ разность фаз между крайними синусоидами будет $\Delta\omega \cdot \tau/2$. Если она сделается равной 2π, то в точке $t = \tau/2$ наложатся синусоиды со всевозможными фазами, непрерывно и равномерно заполняющими интервал шириной 2π. При таком наложении, как легко уяснить с помощью векторной диаграммы, синусоиды погасят друг друга. При том же условии произойдет взаимное гашение синусоид и в точке $t = -\tau/2$. Таким образом, выделится интервал времени $(-\tau/2, +\tau/2)$, на обоих концах которого волновое поле обратится в нуль. На этом участке время τ связано с шириной спектра $\Delta\omega$ соотношением $\Delta\omega \cdot \tau = 4\pi$, которое по существу совпадает с (29.8).

4. Формула, аналогичная (29.8), имеет место и для пространственного распределения волнового поля, когда оно рассматривается в какой-либо фиксированный момент времени. Примером может служить «оборванная синусоида» на рис. 130 *а*, если по оси абсцисс откладывать координату x в направлении распространения волны. При таком рассмотрении снова можно пользоваться формулой (29.4), заменив в ней время t на координату x, а частоту ω на волновое число $k = \omega/v$. Поступив так, легко найти, что длина цуга волн Δx связана с соответствующим интервалом волновых чисел Δk соотношением

$$\Delta k \cdot \Delta x \gtrsim 2\pi. \tag{29.9}$$

Этот результат будет обобщен в § 44 на случай трех измерений. Соотношения (29.8) и (29.9) играют важную роль в квантовой механике, где они при несколько иной интерпретации выражают так называемый *принцип неопределенности Гайзенберга* (1901–1976).

5. Когда ширина спектра $\Delta\omega$ достаточно мала ($\Delta\omega \ll \omega$), то соответствующий свет называют *квазимонохроматическим*. Часто пользуются также *«спектроскопическим волновым числом»* $\tilde{k} = k/(2\pi)$. Для него

$$\Delta\tilde{k} \cdot \Delta x \gtrsim 1. \tag{29.9a}$$

Примером могут служить спектральные линии, излучаемые разреженными газами. Квазимонохроматический свет можно выделить из непрерывного спектра излучения (например, Солнца) призмой, дифракционной решеткой или другими приборами, осуществляющими спектральное разложение. За меру монохроматичности света можно принять отношение $|\omega/\Delta\omega|$ или равное ему отношение $|\lambda/\Delta\lambda|$. Наибольшей монохроматичностью обладает свет, излучаемый газовыми лазерами. Нельзя, однако, упускать из виду, что цуг волн или световой пучок, занимающий спектральную область $\Delta\omega$, должен обладать конечной длительностью, не меньшей $\tau \approx 2\pi/\Delta\omega$. В частности, не имеет смысла говорить о мгновенной объемной плотности энергии излучения в интервале $d\omega$, т.е. о величине $\sim |dE|^2 = |a(\omega)\,d\omega|^2$. Вместо мгновенной надо пользоваться средней плотностью энергии излучения за время порядка $\tau \approx 2\pi/d\omega$. Это значит, что интенсивность излучения в спектральном интервале $(\omega, \omega + d\omega)$ должна определяться величиной, пропорциональной $\tau|dE|^2$, т.е. $|a(\omega)|^2\,d\omega$. Величину $|a(\omega)|^2$ называют *спектральной плотностью излучения.* Именно такую величину имеют в виду, говоря о распределении интенсивности или энергии излучения в спектре. Вместо частоты можно, конечно, пользоваться длинами волн, представляя ту же величину в виде $|a_\lambda(\lambda)|^2\,d\lambda$. Очевидно, $a(\omega) = 2\pi c a_\lambda(\lambda)/\omega^2$.

ЗАДАЧА

Разложить в интеграл Фурье и найти спектральную плотность излучения для затухающего осциллятора, волновое поле которого определяется выражением

$$E(t) = e^{-t/\tau}\sin\omega_0 t \qquad (t \geqslant 0).$$

Ответ. $a(\omega) = \dfrac{1}{2\pi}\dfrac{1}{(\omega-\omega_0) - t/\tau}$,

$$|a(\omega)|^2 = \frac{1}{4\pi^2}\frac{1}{(\omega-\omega_0)^2 + (1/\tau)^2}. \qquad (29.10)$$

Если $\omega - \omega_0 = 1/\tau$, т.е. $\Delta\omega\cdot\tau = 1$, то спектральная плотность излучения уменьшается в два раза.

Этот пример снова подтверждает общее соотношение (29.8). В примере, рассмотренном в тексте (см. рис. 130 *б*), спектральная плотность излучения $|a(\omega)|^2$ на краях интервала $-\dfrac{\pi}{2} < \alpha < \dfrac{\pi}{2}$ убывает в $(\pi/2)^2 \approx 2{,}5$ раза по сравнению с той же величиной при $\alpha = 0$.

§ 30. Влияние немонохроматичности света

1. Как и увеличение размеров источников, немонохроматичность света ведет сначала к *ухудшению контрастности* (видимости) интерференционных полос, а затем к полному исчезновению их. Чтобы не усложнять исследование учетом конечных размеров источника, будем предполагать, что источник света S точечный. Пусть S_1 и S_2 (см. рис. 113) — когерентные источники, являющиеся действительными или мнимыми изображениями источника S. Допустим сначала, что излучение источника S состоит из двух близких одинаково интенсивных спектральных линий с длинами волн λ и $\lambda' = \lambda + \delta\lambda$. Точка или линия экрана, где оптическая разность хода Δ интерферирующих лучей равна нулю, называется *центром интерференционной картины.*

Если начальные фазы источников S_1 и S_2 одинаковы, то в центр картины лучи с длинами волн λ и λ' придут в *одинаковых фазах.* Для обеих волн там получится светлая полоса. В другой точке экрана A, в которой $\Delta = N\lambda'$, где N — целое число (номер полосы или порядок интерференции), для длины волны λ' получится также светлая интерференционная полоса. Если $\Delta = (N + {}^1\!/\!_2)\lambda$, то в ту же точку A интерферирующие лучи с другой длиной волны λ придут уже *в противоположных фазах*, и для такой длины волны интерференционная полоса будет темной. При этом условии в окрестности точки A светлые полосы с длиной волны λ' наложатся на темные полосы с длиной волны λ. Интерференционные полосы в указанной окрестности исчезнут. Условие первого исчезновения полос, таким образом, есть $N\lambda' = (N + {}^1\!/_2)\lambda$, или

$$N = \frac{\lambda}{2(\lambda' - \lambda)} = \frac{\lambda}{2\,\delta\lambda}. \tag{30.1}$$

Все изложенное остается верным и в том случае, когда фазы лучей, приходящих в центр интерференционной картины, противоположны. Только в этом случае центральная полоса будет темной.

Когда номер полосы мал по сравнению с величиной N, определяемой выражением (30.1), интерференционные полосы будут почти столь же отчетливы, что и в случае света с одной длиной волны. Когда номер полосы для длины волны λ' достигнет значения $2N$, номер соответствующей полосы для длины волны λ' сделается равным $(2N + 1)$. Тогда полосы интерференции сделаются столь же резкими, что и в центре интерференционной картины. При дальнейшем возрастании порядка интерференции будет наблюдаться периодическая смена резкости интерференционных полос от наибольшей отчетливости их до полного исчезновения.

Например, когда источником света является пламя натрия, то излучаются две узкие спектральные линии с длинами волн $\lambda = 589$ и $\lambda' = 589{,}6$ нм. При $N \approx \lambda/[2(\lambda' - \lambda)] = 490$ интерференционные полосы становятся очень неясными или совсем ненаблюдаемыми. При N, равном 980 или кратном этому числу, полосы опять становятся отчетливыми. При дальнейшем возрастании порядка интерференции полосы периодически меняют свой вид, становясь попеременно то отчетливыми, то размытыми. И так продолжается до тех пор, пока из-за конечной ширины обеих спектральных линий интерференционная картина не пропадет совсем.

2. Перейдем теперь к случаю, когда свет от источника S непрерывно и равномерно заполняет спектральный интервал $(\lambda, \lambda + \delta\lambda)$. В этом случае можно поступить так же, как мы поступали в § 28 (п. 2) при рассмотрении интерференционных полос от протяженного источника света. Весь спектральный интервал $(\lambda, \lambda + \delta\lambda)$ разобьем на множество пар бесконечно узких спектральных линий, находящихся на расстоянии $\delta\lambda/2$ друг от друга (на шкале длин волн). К каждой такой паре применима формула (30.1), если в ней сделать замену $\delta\lambda \to \delta\lambda/2$. Поэтому первое исчезновение интерференционных полос произойдет для порядка интерференции

$$N = \lambda/\delta\lambda, \tag{30.2}$$

вдвое большего, чем в предыдущем случае. При дальнейшем возрастании порядка интерференции интерференционные полосы появятся вновь, но уже на светлом фоне, аналогично тому, как это было в случае протяженного монохроматического источника света (§ 28, п. 2, а также задача 3). Их контрастность будет незначительна, а при дальнейшем возрастании порядка интерференции и совсем исчезнет. Практически наблюдение интерференционных полос ограничится порядками интерференции, не превосходящими (30.2).

Формула (30.2) как оценочная остается верной и в случае произвольного распределения интенсивности света по длинам волн в интервале $\delta\lambda$. Таким образом, эта формула дает оценку максимально возможного порядка интерференции при заданной степени монохроматичности $\lambda/\delta\lambda$ используемого света. Два квазимонохроматических пучка света, когерентные при низких порядках интерференции, перестают быть таковыми при высоких порядках, превышающих примерно величину (30.2).

3. Максимально возможный порядок интерференции можно также найти из следующих соображений. Пусть используемый свет состоит из одинаковых цугов волн, следующих друг за другом через беспорядочно меняющиеся промежутки времени. Один из возможных цугов изображен на рис. 130 *а* (здесь показано колебание волнового поля во времени в фиксированной точке пространства; совершенно аналогично выглядит цуг в пространстве в фиксированный момент времени). Интерференционный прибор разделяет пучок света на два пучка, идущие к месту схождения по разным путям, так что между пучками возникает разность хода. Для когерентности пучков необходимо, чтобы разность хода между ними не превосходила длину цуга $L = c\tau$. В противоположном случае будет происходить наложение независимых цугов волн, испущенных в разные моменты времени, и интерференция не возникнет. Максимальный порядок интерференции не может превышать величину

$$N_{\text{макс}} = \frac{L}{\lambda} = \frac{\tau}{T}. \tag{30.3}$$

Разумеется, это условие должно согласовываться с условием (30.2), а потому должно быть

$$\frac{L}{\lambda} = \frac{\tau}{T} = \frac{\lambda}{\delta\lambda} = \frac{\omega}{\delta\omega}.$$

Если учесть, что $\lambda = 2\pi/k$ и $T = 2\pi/\omega$, то отсюда получится

$$\tau \cdot \delta\omega = 2\pi \quad \text{и} \quad L \cdot \delta k = 2\pi.$$

К таким соотношениям мы уже пришли в предыдущем параграфе из других соображений.

С изложенной точки зрения нарушения когерентности связаны с *запаздыванием* одного цуга волн по сравнению с другим. Поэтому здесь, в отличие от пространственной, говорят о *временной когерентности*, а длительность цуга τ называют *временем когерентности*. Временная когерентность — это то же самое, что и когерентность, связанная с *узостью спектрального интервала* $\Delta\omega$, занимаемого светом. В силу (29.8) время когерентности $\tau_{\text{ког}}$ связано с шириной спектрального интервала $\Delta\omega$ соотношением

$$\tau_{\text{ког}} \approx \frac{2\pi}{\Delta\omega} \approx \frac{1}{\Delta\nu}. \tag{30.4}$$

Чем меньше $\Delta\nu$, тем больше время когерентности и тем выше порядок интерференции, который может наблюдаться при заданном ν. Максимальная разность хода, при которой еще возможна интерференция, определяется соотношением

$$L \approx c\tau_{\text{ког}} = \lambda\frac{\nu}{\delta\nu} = \frac{\lambda^2}{\delta\lambda}.$$

Она называется *длиной когерентности*. Ей соответствует максимальный порядок интерференции

$$N_{\text{макс}} = \frac{\nu}{\delta\nu} = \frac{\lambda}{\delta\lambda}.$$

Значения параметров $\delta\nu$, τ и L очень резко различаются для света тепловых источников и света, генерируемого газовыми лазерами. Ширина спектральной линии $\delta\nu$ лучших «монохроматических» тепловых источников, которые могут быть созданы в лаборатории, порядка 10^8 Гц, тогда как в случае лазеров можно получить $\delta\nu \sim 10^2$ Гц или даже еще меньше. Соответствующие времена когерентности будут 10^{-8} и 10^{-2} с, а длины когерентности 1 и 10^6 м. Лазерные источники света позволяют наблюдать интерференцию при разности хода в несколько километров. Здесь максимальный порядок интерференции, который можно наблюдать, ограничивается не степенью монохроматичности лазерного излучения, а неоднородностью земной атмосферы и трудностями создания стабильной интерференционной схемы столь больших размеров.

4. Для белого света $\delta\lambda \sim \lambda$, т.е. $N \sim 1$. Казалось бы, что в белом свете интерференционные полосы наблюдаться не должны. Это действительно так, если пользоваться такими приемниками света, как фотоэлемент, болометр или термостолбик, которые обладают примерно одинаковой чувствительностью в различных участках спектра. Но глаз — *селективный приемник*, т.е. его чувствительность к различным длинам волн разная (см.

кривую видности человеческого глаза на рис. 82). Именно поэтому в белом свете глаз видит около десятка интерференционных полос. Полосы цветные, так как из-за различия в длинах волн полосы разного цвета имеют разную ширину и сдвинуты относительно друг друга. Только в центре картины, куда волны от обоих источников приходят в одинаковых фазах, соблюдается условие максимума для всех длин волн. Там получается *ахроматическая*, т.е. неокрашенная, светлая интерференционная полоса. В опыте с зеркалом Ллойда ахроматическая полоса темная. Это доказывает, что отражение света от зеркала сопровождается изменением фазы волны на π (см. § 33, 65). Конечно, при обычной постановке опыта (см. рис. 119) ахроматическую полосу увидеть нельзя, так как луч, отражающийся от зеркала, проходит больший путь, чем прямой луч. Чтобы получить эту полосу, надо на пути прямого луча поставить плоскопараллельную пластинку, смещающую вверх всю интерференционную картину.

§ 31. Корреляция и когерентность света

1. При точном количественном определении понятия когерентности надо учесть, что реальные световые колебания *не синусоидальны*. Напряженность поля в каждой точке пространства может быть представлена интегралом Фурье (29.4), т.е. в виде суперпозиции синусоидальных колебаний различных частот. Если область $\Delta\omega$, заполняемая этими частотами, мала по сравнению с самими частотами ω, входящими в суперпозицию, то результирующее колебание и представляемый им свет называются *квазимонохроматическими*. Выбрав внутри интервала $\Delta\omega$ произвольную частоту ω_0, запишем квазимонохроматическое колебание в виде

$$E(t) = a(t)e^{i\omega_0 t}, \qquad (31.1)$$

где $a(t)$ — комплексная амплитуда, медленно меняющаяся по сравнению с быстро осциллирующей функцией $e^{i\omega_0 t}$. Она, конечно, определена не совсем однозначно, поскольку ее значение зависит от выбора частоты ω_0. Про колебание, представленное формулой (31.1), говорят, что оно *модулировано*.

При модуляции может медленно меняться (вещественная) амплитуда, колебания или его (начальная) фаза. В первом случае говорят об *амплитудной*, во втором — о *фазовой модуляции*. Могут также одновременно меняться и амплитуда, и фаза. В случае квазимонохроматического света, излучаемого реальными источниками, такие изменения происходят хаотически — амплитуда и фаза являются *случайными функциями времени*. Поэтому при изучении реального света, в том числе и квазимонохроматического, нельзя обойтись без использования *статистических методов*. Для простоты мы отвлечемся от векторного характера колебаний, считая их скалярными.

Важно отметить, что из-за очень высоких оптических частот все существующие приемники света не позволяют регистрировать быстрые изменения напряженности световых полей за времена порядка периода световых колебаний. Обычно они не позволяют следить и за быстрыми изменениями световых потоков, обусловленных случайными изменениями амплитуд и фаз колебаний. Удается измерять только *квадраты напряженностей* световых полей, усредненные по промежуткам времени, весьма большим не только по сравнению с периодами световых колебаний, но и по сравнению с временами, в течение которых происходят случайные изменения амплитуд и фаз этих колебаний. Ниже предполагается, что световые потоки регистрируются именно такими «инерционными» приемниками.

Более того, мы будем предполагать, что световые потоки в среднем *стационарны*, т.е. значение среднего квадрата поля в каждой точке пространства одинаково для всех моментов времени и не зависит от положения на шкале времени временного интервала, по которому производится усреднение. Квадрат поля можно представить в виде

$$(\mathrm{Re}\,E)^2 = \left(\frac{E+E^*}{2}\right)^2 = \frac{1}{4}(E^2+E^{*2}) + \frac{1}{2}EE^*.$$

Если положить $a = a_0(t)e^{i\delta(t)}$, где $a_0(t)$ и $\delta(t)$ — *медленно меняющиеся* вещественная амплитуда и фаза, то

$$E^2 + E^{*2} = a_0^2[e^{2i(\omega_0 t+\delta)} + e^{-2i(\omega_0 t+\delta)}] = 2a_0^2\cos[2(\omega_0 t+\delta)].$$

Эта величина осциллирует во времени очень быстро и при усреднении пропадает. Результат усреднения величины $(\mathrm{Re}\,E)^2$ определяется только последним членом $^1\!/_2 EE^*$. Поэтому за меру интенсивности колебаний можно принять величину $\overline{EE^*}$.

2. Допустим теперь, что в точку наблюдения P в момент времени t приходят два колебания от источников света S_1 и S_2 (см. рис. 113). Чтобы прийти в P в момент t, эти колебания должны выйти из S_1 и S_2 в более ранние моменты времени $t-\theta_1$ и $t-\theta_2$, где θ_1 и θ_2 — времена, затрачиваемые светом на распространение от S_1 и S_2 до точки P. Чтобы отметить это, рассматриваемые колебания в точке P обозначим через $E_1(t-\theta_1)$ и $E_2(t-\theta_2)$ соответственно. При их сложении в точке P получится результирующее колебание

$$E \equiv E(P,t) = E_1(t-\theta_1) + E_2(t-\theta_2).$$

Для нахождения его интенсивности в точке P умножим это равенство на комплексно-сопряженное и произведем усреднение по времени. В результате получим

$$I = \overline{E_1(t-\theta_1)E_1^*(t-\theta_1)} + \overline{E_2(t-\theta_2)E_2^*(t-\theta_2)} + \\ + \overline{E_1(t-\theta_1)E_2^*(t-\theta_2) + E_1^*(t-\theta_1)E_2(t-\theta_2)}.$$

В силу предположения о стационарности (в среднем) световых потоков первое слагаемое справа не зависит от θ_1 и t. Оно представляет просто интенсивность I_1 первого колебания, пришедшего в точку P:

$$I_1 = \frac{1}{\tau}\int\limits_{-\tau/2}^{\tau/2} E_1(t-\theta_1)E_1^*(t-\theta_1)\,dt = \frac{1}{\tau}\int\limits_{-\tau/2}^{\tau/2} E_1(t)E_1^*(t)\,dt,$$

где τ — ширина временного интервала, по которому производится усреднение. Аналогично, второе слагаемое есть интенсивность I_2 второго колебания. По той же причине последнее слагаемое (интерференционный член) не зависит от t, а также от θ_1 и θ_2 в отдельности. Оно есть функция только разности $\theta = \theta_2 - \theta_1$, т.е. времени запаздывания второго колебания относительно первого. Поэтому можно положить

$$\overline{E_1(t-\theta_1)E_2^*(t-\theta_2)} = \overline{E_1(t)E_2^*(t-\theta)} = F_{12}(\theta), \tag{31.2}$$

где $F_{12}(\theta)$ — комплексная функция, характеризующая *степень согласованности* рассматриваемых колебаний в точке P. Она называется *корреляционной функцией* колебаний $E_1(t-\theta_1)$ и $E_2(t-\theta_2)$ или *взаимной корреляционной функцией.*

В частном случае функции $E_1(t)$ и $E_2(t)$ могут оказаться *тождественными.* Это будет, например, когда оба колебания выходят из одного и того же источника, но приходят в точку P по различным путям. Тогда $F_{12}(\theta)$ называют *автокорреляционной функцией.* Ее можно обозначать через $F_{11}(\theta)$, но мы чаще будем применять обозначение $F(\theta)$. При $\theta = 0$ автокорреляционная функция переходит в $\overline{|E_1(t-\theta_1)|^2}$, т.е. в интенсивность колебания I_1 в точке P.

Функция $F_{12}(\theta)$ зависит от I_1 и I_2, т.е. от интенсивностей складываемых колебаний в точке P. Если положить

$$F_{12}(\theta) = \sqrt{I_1 I_2}\, f_{12}(\theta), \tag{31.3}$$

то получится *нормированная корреляционная функция,* которая зависит только от времени запаздывания θ, но уже не зависит от I_1 и I_2. Через эту функцию результирующая интенсивность в точке P представляется выражением

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\,\mathrm{Re}\,[f_{12}(\theta)]. \tag{31.4}$$

Для квазимонохроматического света $E_1(t) = a_1(t)e^{i\omega_0 t}$, $E_2(t) = a_2(t)e^{i\omega_0 t}$, так что

$$\overline{a_1(t)a_2^*(t-\theta)}e^{i\omega_0\theta} = \sqrt{I_1 I_2}\, f_{12}(\theta). \tag{31.5}$$

Как видно из этой формулы, величина $f_{12}(\theta)$ есть быстро меняющаяся функция времени запаздывания θ. Разделив ее на столь же быстро меняющуюся осциллирующую функцию $e^{i\omega_0\theta}$, получим уже медленно меняющуюся функцию

$$\gamma_{12}(\theta) = f_{12}(\theta)e^{-i\omega_0\theta}, \tag{31.6}$$

которая называется *комплексной степенью когерентности колебаний,* а ее модуль $|\gamma_{12}(\theta)| = |f_{12}(\theta)|$ — просто *степенью когерентности колебаний* в точке P. Таким образом,

$$\overline{a_1(t)a_2^*(t-\theta)} = \sqrt{I_1 I_2}\,\gamma_{12}(\theta), \tag{31.7}$$

т.е. $\gamma_{12}(\theta)$ есть *нормированная взаимная корреляционная функция для амплитуд* $a_1(t)$ *и* $a_2(t)$. Далее,

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\,\mathrm{Re}\,[\gamma(\theta)e^{i\omega_0\theta}]. \tag{31.8}$$

Полагая $\gamma_{12}(\theta) = |\gamma_{12}(\theta)|e^{i\delta}$, запишем последний результат в вещественной форме:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\,|\gamma_{12}(\theta)|\cos(\omega_0\theta + \delta). \tag{31.9}$$

Эта формула отличается от аналогичной формулы (26.7) для строго синусоидальных колебаний добавочным множителем $|\gamma_{12}(\theta)|$ в интерференционном члене и добавочным, медленно меняющимся слагаемым $\delta(\theta)$ в разности фаз. Значения вещественных амплитуд $|a_1|$ и $|a_2|$ и соответствующих интенсивностей I_1 и I_2 не зависят от выбора промежуточной частоты ω_0 в спектральном интервале $\Delta\omega$ квазимонохроматического света. Не может зависеть от выбора ω_0 и полная фаза $\omega_0\theta + \delta$, входящая в формулу (31.9). Но добавочная фаза δ, конечно, будет другой при другом выборе ω_0. Фаза $\omega_0\theta + \delta$ определяет наиболее быстрые изменения в пространстве интенсивности светового поля, т.е. изменения при переходе от одной интерференционной полосы к другой. Ввиду медленности изменения функции $|\gamma_{12}(\theta)|$, ее изменениями при таком переходе можно пренебречь. Тогда в максимумах $\cos(\omega_0\theta + \delta)$ будет равен $+1$, а в минимумах -1. Поэтому

$$I_{\text{макс}} = I_1 + I_2 + 2\sqrt{I_1 I_2}\,|\gamma_{12}(\theta)|, \qquad I_{\text{мин}} = I_1 + I_2 - 2\sqrt{I_1 I_2}\,|\gamma_{12}(\theta)|.$$

Отсюда для видности интерференционных полос находим

$$V \equiv \frac{I_{\text{макс}} - I_{\text{мин}}}{I_{\text{макс}} + I_{\text{мин}}} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}|\gamma_{12}(\theta)|. \qquad (31.10)$$

Когда интенсивности складываемых колебаний одинаковы ($I_1 = I_2$), то $V = |\gamma_{12}(\theta)|$. По самому определению видность V не может быть больше единицы, а функция $\gamma_{12}(\theta)$ от интенсивностей пучков не зависит. Поэтому всегда $|\gamma_{12}(\theta)| \lesssim 1$. Когда $|\gamma_{12}(\theta)| = 0$, то $V = 0$, т.е. интерференционных полос не получается. В этом случае колебания называются *некогерентными*. Если при этом функция $\gamma_{12}(\theta)$ обращается в нуль при любых значениях θ, то некогерентность называется *полной*. Тогда всюду $I = I_1 + I_2$, т.е. имеет место *закон фотометрического сложения* интенсивностей. Такой случай осуществляется при наложении световых пучков от независимых источников света.

Если же $\gamma_{12}(\theta) \neq 0$, то наблюдается интерференция, и колебания называются *когерентными*. Когерентность называется *полной*, когда величина $|\gamma_{12}(\theta)|$ всюду достигает своего предельного значения 1. В этом случае интерференционные полосы наиболее контрастны, т.е. при заданных I_1 и I_2 видность V максимальна. Такой случай реализуется при наложении строго периодических, в частности монохроматических, пучков одинаковых периодов. Во всех остальных случаях (когда $0 < |\gamma_{12}(\theta)| < 1$) говорят о *частичной когерентности*. При перемещении точки наблюдения степень когерентности $|\gamma_{12}(\theta)|$ медленно изменяется. Вследствие этого медленно изменяется и видность интерференционных полос.

3. До сих пор речь шла о когерентности двух колебаний, происходящих в *одной и той же точке* пространства. Но можно говорить о когерентности *одного и того же волнового поля* в двух различных пространственно-временных точках $R_1(Q_1, t_1)$ и $R_2(Q_2, t_2)$. Этот вопрос сводится к предыдущему.

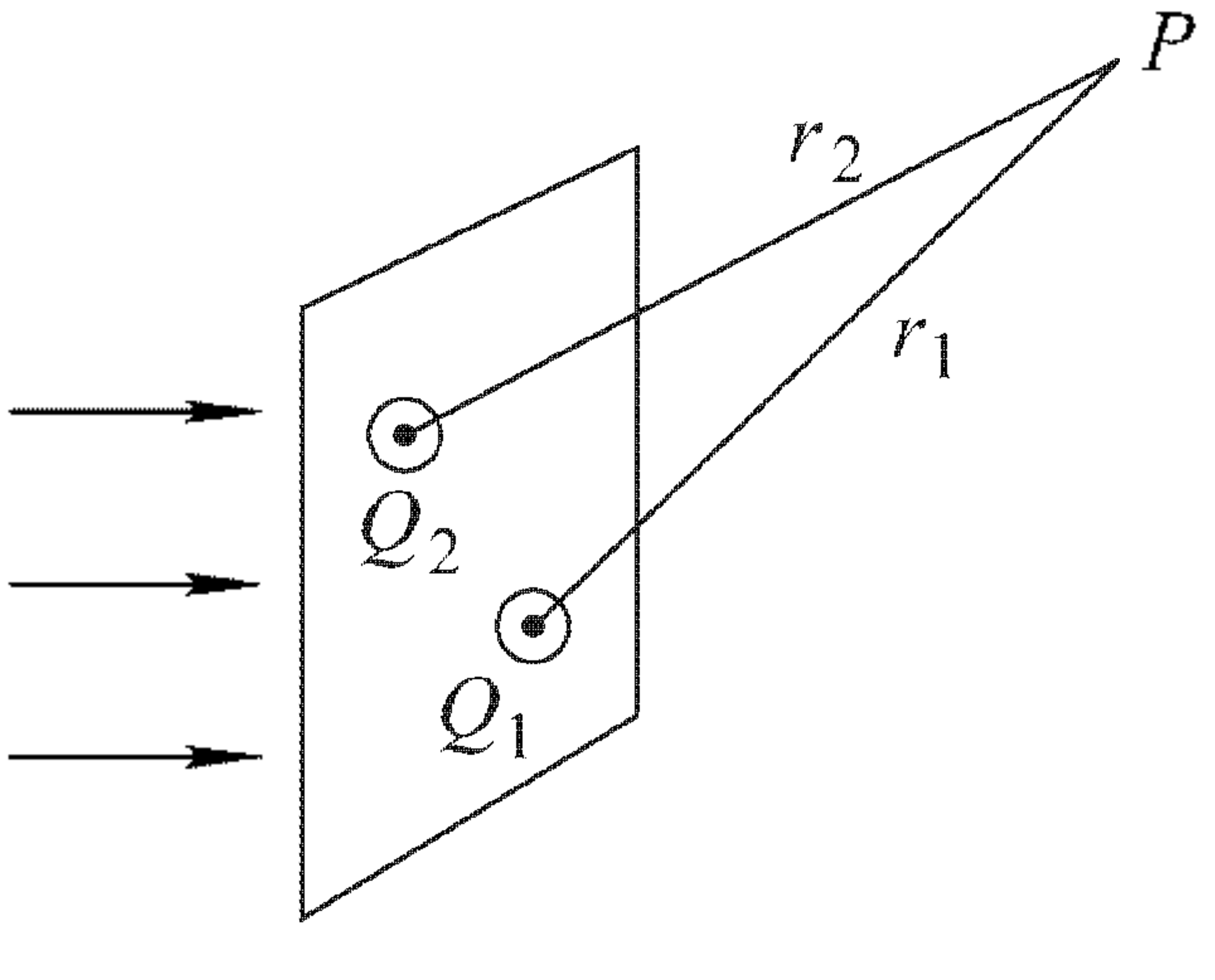


Рис. 131

Пусть Q_1 и Q_2 — какие-либо две точки пространства, находящиеся в рассматриваемом поле излучения (рис. 131). Пусть они являются центрами двух бесконечно малых отверстий в непрозрачном экране, поставленном на пути распространения света. Экран всюду загородит падающий свет, но пропустит свет через отверстия. Через отверстия пройдет не только прямой, но и *дифрагированный свет.* Бесконечно малые отверстия в силу принципа Гюйгенса могут рассматриваться как точечные *вторичные источники*, посылающие свет за экран во всех направлениях. Возьмем за экраном удаленную точку наблюдения P. Пусть колебания, вышедшие из точек Q_1 и Q_2 в моменты времени t_1 и t_2, приходят в точку P *одновременно.* Тогда можно говорить о когерентности этих колебаний в том смысле, как это было разъяснено выше.

По определению мы называем колебания в точках Q_1 и Q_2 в моменты времени t_1 и t_2 (т.е. в пространственно-временных точках R_1 и R_2) *когерентными* или *некогерентными*, если когерентны или некогерентны соответствующие колебания в точке P. При этом степень когерентности $\gamma(\theta)$ мы определяем той же величиной, что и для колебаний в точке наблюдения.

В частности, если пространственные точки Q_1 и Q_2 совпадают, но свет попадает в P различными путями, то пространственно-временные точки $R_1(Q_1, t_1)$ и $R_2(Q_2, t_2)$ отличаются только моментами времени t_1 и t_2. В этом случае говорят о *временной когерентности.* При $t_1 = t_2$ степень временной когерентности равна единице. С увеличением разности этих времен степень когерентности убывает. Максимальное значение $|t_1 - t_2|$, при котором когерентность еще сохраняется, называется *временем когерентности.* Расстояние $v|t_1 - t_2|$, проходимое светом эа это время, называется *длиной когерентности.*

В другом крайнем случае времена t_1 и t_2 одинаковы, но пространственные точки Q_1 и Q_2 не совпадают. Тогда говорят о *пространственной когерентности.* Сохраняя точку Q_1 неподвижной, будем поворачивать вокруг нее экран вместе с точкой Q_2. Тогда точка Q_2 будет перемещаться вокруг Q_1, а степень когерентности $|\gamma_{12}|$ будет меняться. Геометрическое место точек, где γ_{12} обращается в нуль, есть некоторая поверхность, окружающая точку Q_1. Объем, который она ограничивает, называется *объемом когерентности* вокруг точки P_1.

Вычисление степени временной когерентности может быть систематически использовано при определении допустимой ширины спектральной области, а пространственной когерентности — допустимых размеров источников света для возможности наблюдения интерференции.

4. Иллюстрируем понятие и свойства автокорреляционной функции и степени когерентности на простейшем примере, когда оба колебания представляются «оборванной синусоидой»: $E(t) = \sin\omega_0 t$ в интервале $0 < t < \tau$ и $E(t) = 0$ вне этого интервала. Перейдем к комплексной форме $E(t) = \frac{1}{i}e^{i\omega_0 t}$ и примем τ за промежуток времени, по которому производится усреднение. Тогда $I_1 = I_2 = 1$. Произведение $E(t)E^*(t-\theta)$ отлично от нуля только в интервале $\theta < t < \tau$. Поэтому только при $\theta < \tau$ функция $\gamma(\theta)$ может отличаться от нуля. При $\theta > \tau$ она обращается в нуль. В первом случае

$$\overline{E(t)E^*(t-\theta)} = \frac{1}{\tau}\int\limits_0^\tau e^{i\omega_0\theta}\,dt = \frac{\tau-\theta}{\tau}e^{i\omega_0\theta} = F(\theta) = f(\theta).$$

Следовательно,

$$\gamma(\theta) = \begin{cases} 1 - \theta/\tau & \text{при} \quad \theta < \tau, \\ 0 & \text{при} \quad \theta > \tau. \end{cases} \tag{31.11}$$

К тому же результату мы пришли бы, если бы предположили, что источник света излучает цуги волн одинаковой длительности τ, беспорядочно следующие друг за другом, причем каждый цуг разделяется на две части, идущие к точке наблюдения различными путями. Это непосредственно следует из того, что различные цуги, испускаемые источником, *статистически независимы* и поэтому не интерферируют между собой. Из формулы (31.11) следует физически очевидный результат, что колебания когерентны, если время запаздывания θ меньше длины цуга τ. В противоположном случае они некогерентны. Значит, τ есть время когерентности колебаний.

5. Модуль функции $|\gamma_{12}(\theta)|$ легко вычислить по формуле (31.10), измерив предварительно видность полос V и интенсивности I_1 и I_2 накладывающихся пучков в точке наблюдения. Значительно труднее измерить добавочную фазу δ, входящую в формулу (31.9). Особенно трудно это сделать, когда источниками света являются узкие спектральные линии. Для этого надо сравнить в *одном и том же месте* интерференционной картины номера интерференционных полос от рассматриваемого источника света с номерами полос от источника с частотой ω_0. Для номера максимума N-й интерференционной полосы от первого источника можно написать $\omega_0\theta+\delta = 2\pi N$. В том же месте второй источник, вообще говоря, не даст максимума. Этому месту будет соответствовать уже дробное число интерференционных полос, определяемое условием $\omega_0\theta = 2\pi N_0$. Отсюда $\delta = 2\pi(N - N_0)$. Таким путем в принципе можно экспериментально определить не только модуль, но и аргумент комплексной степени когерентности $\gamma_{12}(\theta)$. Вместе с тем можно определить и корреляционную функцию $F_{12}(\theta)$.

6. Автокорреляционная функция $F(\theta)$ связана важным соотношением со спектральной плотностью $I_\omega(\omega)$ излучения. Для установления этой связи пишем на основании определения автокорреляционной функции:

$$F(\theta) = \overline{E(t)E^*(t-\theta)} = \frac{1}{\tau}\int\limits_{-\tau/2}^{\tau/2} E(t)E^*(t-\theta)\,dt. \tag{31.12}$$

Подставим сюда

$$E^*(t-\theta) = \int\limits_0^\infty a^*(\omega)e^{-i\omega(t-\theta)}\,d\omega$$

и поменяем порядок интегрирования по t и ω. Использовав при этом формулу (29.5), получим

$$F(\theta) = \frac{2\pi}{\tau}\int\limits_0^\infty a^*(\omega)a(\omega)e^{i\omega\theta}\,d\omega.$$

Но $(2\pi/\tau)a^*(\omega)a(\omega)$ есть спектральная плотность излучения $I_\omega(\omega)$ (см. § 29, п. 5). Следовательно,

$$F(\theta) = \int\limits_0^\infty I_\omega(\omega)e^{i\omega\theta}\,d\omega. \tag{31.13}$$

Эта формула представляет фурье-разложение функции $F(\theta)$, а потому

$$I_\omega(\omega) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} F(\theta)e^{-i\omega\theta}\,d\theta. \tag{31.14}$$

Формулу (31.14) можно привести к другому виду. Для этого заметим, что, ввиду стационарности светового потока, в формуле (31.12) пределы интегрирования можно заменить любыми другими, сохраняя только неизменной ширину интервала интегрирования. Используя это, нетрудно доказать, что автокорреляционная функция удовлетворяет соотношению $F(-\theta) = F^*(\theta)$. После этого формула (31.14) приводится к виду

$$I_\omega(\omega) = \frac{1}{2\pi}\left[\int\limits_0^\infty F(\theta)e^{-i\omega\theta}\,d\theta + \int\limits_0^\infty F^*(\theta)e^{i\omega\theta}\,d\theta\right]. \tag{31.15}$$

Это соотношение можно записать в символическом виде

$$I_\omega(\omega) = \frac{1}{\pi}\int\limits_0^\infty F(\theta)e^{-i\omega\theta}\,d\theta, \tag{31.16}$$

понимая его в том смысле, что левая часть равенства равна вещественной части правой. Соотношение (31.13) позволяет найти корреляционную функцию $F(\theta)$ по экспериментально измеренной спектральной плотности излучения $I_\omega(\omega)$. С помощью обратного соотношения (31.16) можно определить спектральную плотность $I_\omega(\omega)$, если экспериментально определить корреляционную функцию $F(\theta)$.

Именно так поступал Майкельсон, используя свой интерферометр для исследования структуры спектральных линий. Он измерял видность интерференционных полос в интерферометре и фазу δ, входящую в формулу (31.9), и по этим данным вычислял спектральную плотность излучения $I_\omega(\omega)$. В свое время из всех методов этот метод был наиболее точным. Позднее метод Майкельсона был вытеснен более простыми методами многолучевой интерферометрии.

§ 32. Теорема Ван-Циттера–Цернике

Определим комплексную степень пространственной когерентности γ_{12} для точек Q_1 и Q_2 экрана Э, освещаемого протяженным квазимонохроматическим самосветящимся источником света (рис. 132). Сами точки Q_1 и Q_2 должны рассматриваться как вторичные источники волн в том смысле, как это было указано в п. 3 предыдущего параграфа. Будем предполагать, что точка наблюдения P равноудалена от Q_1 и Q_2. При этом условии вместо волн, накладывающихся в точке P, можно брать волновые поля, создаваемые первичным источником в точках Q_1 и Q_2, что и будет делаться в дальнейшем. Наши вычисления, следовательно, проводятся в предположении, что время запаздывания θ равно нулю. Однако, ввиду медленности изменения функции $\gamma_{12}(\theta)$, вычисленным значением ее можно пользоваться не только при $\theta = 0$, но и при малых значениях аргумента θ, пока видность интерференционных полос не претерпит существенных изменений.

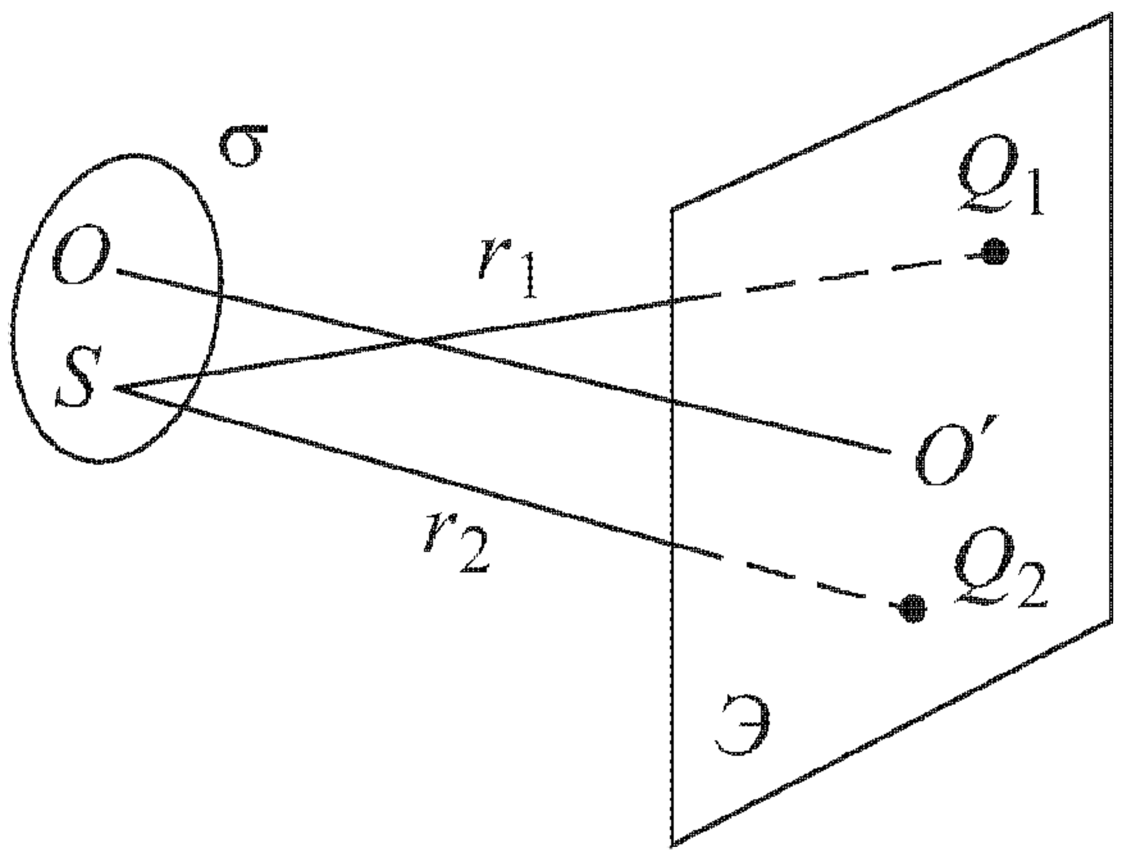


Рис. 132

Для простоты в качестве источника света возьмем малую площадку σ, плоскость которой параллельна плоскости экрана Э. Среда между источником σ и экраном Э предполагается однородной, а скорость света в ней обозначается через v. Линейные размеры площадки σ должны быть малы по сравнению с расстоянием ее до экрана. Предполагается также, что малы углы между «средней линией» OO' и прямыми, соединяющими произвольную точку S источника с точками Q_1 и Q_2.

Разобьем источник σ на малые площадки, линейные размеры которых малы по сравнению с длиной волны λ. Будем рассматривать их как некогерентные точечные источники, излучающие сферические волны. Волновые поля, создаваемые таким источником в точках Q_1 и Q_2, представятся выражениями

$$\frac{A_m(t-r_{1m}/v)}{r_{1m}}\, e^{i(\omega_0 t-kr_{1m})} \quad \text{и} \quad \frac{A_m(t-r_{2m}/v)}{r_{2m}}\, e^{i(\omega_0 t-kr_{2m})},$$

где r_{1m} и r_{2m} — расстояния m-го источника до точек Q_1 и Q_2 соответственно. Амплитуды результирующих колебаний в точках Q_1 и Q_2 будут

$$a_1(t)=\sum_m \frac{A_m(t-r_{1m}/v)}{r_{1m}}\, e^{-ikr_{1m}},$$
$$a_2(t)=\sum_m \frac{A_m(t-r_{2m}/v)}{r_{2m}}\, e^{-ikr_{2m}}.$$

Вычислим теперь взаимную корреляционную функцию амплитуд a_1 и a_2, т.е. среднее по времени от произведения $a_1(t)a_2^*(t-\theta)$, в предположении, что $\theta=0$. Перемножив почленно предыдущие суммы, заметим, что все слагаемые с различными m при усреднении обратятся в нуль, ввиду статистической независимости соответствующих элементарных источников света. В результате получим

$$a_1a_2^* = \sum_m \frac{\overline{A_m(t-r_{1m}/v)A_m^*(t-r_{2m}/v)}}{r_{1m}r_{2m}}\, e^{ik(r_{2m}-r_{1m})}.$$

(Усреднению подвергается только числитель, так как все прочие величины от времени не зависят.) Если предположить, что для всех m разности $r_{2m}-r_{1m}$ малы по сравнению с длиной когерентности, то различие аргументов $t-r_{1m}/v$ и $t-r_{2m}/v$ можно не учитывать. Следовательно, в силу предполагаемой однородности световых потоков можно опустить и сами аргументы, т.е.

$$\overline{A_m(t-r_{1m}/v)A_m^*(t-r_{2m}/v)} = \overline{A_mA_m^*}.$$

Во всех практически интересных случаях число элементарных излучателей света очень велико, так что их можно считать непрерывно распределенными по площадке σ с определенной поверхностной плотностью. Тогда от суммы можно перейти к интегралу. Если $I(S)$ — интенсивность света, создаваемая единицей площади источника на единичном расстоянии от него, то $\overline{A_mA_m^*}=I(S)\,dS$. В точках Q_1 и Q_2 соответственно

$$I_1\equiv I(Q_1)=\int_\sigma \frac{I(S)\,dS}{r_1^2},\qquad I_2\equiv I(Q_2)=\int_\sigma \frac{I(S)\,dS}{r_2^2}.$$

Введя еще нормирующий множитель $1/\sqrt{I_1I_2}$, получим окончательно

$$\gamma_{12}(0)=\frac{1}{\sqrt{I_1I_2}}\int \frac{I(S)}{r_1r_2}\, e^{ik(r_2-r_1)}\,dS. \tag{32.1}$$

Для наглядной интерпретации полученного результата воспользуемся следующей аналогией. Пусть точка Q_2 неподвижна, а точка Q_1 может занимать различные положения на экране Э. Заменим площадку σ отверстием σ' той же формы в непрозрачном экране. Допустим, что на него падает сферическая волна, сходящаяся в центре Q_2, волновое поле которой в точках отверстия представляется выражением

$$\psi = \frac{I(S)}{\sqrt{I_1 I_2}}\, e^{i(\omega_0 t + kr_2)} \equiv \psi_0(r_2) e^{i\omega_0 t}.$$

Пусть каждый элемент dS площади отверстия излучает по принципу Гюйгенса вторичную сферическую волну, поле которой в точке P_1 определяется выражением $\frac{1}{r_1}\,\psi_0 e^{i(\omega_0 t - kr_1)}$. Тогда в результате суперпозиции таких вторичных волн получится волна, комплексная амплитуда которой определяется формулой (32.1). Но именно так по принципу Гюйгенса решается задача о дифракции сферической волны на отверстии σ' (см. § 39). Следовательно, *комплексная степень взаимной когерентности в точках Q_1 и Q_2 ровна комплексный амплитуде в точке Q_1 соответствующей дифрагированной волны.*

В этом состоит *теорема Ван-Циттера–Цернике.* Она сводит вычисление степени взаимной когерентности γ_{12} к соответствующей задаче дифракции.

§ 33. Интерференция в пленках и пластинках

1. При освещении тонкой пленки происходит наложение волн от одного и того же источника, отразившихся от передней и задней поверхностей пленки. При этом может возникнуть интерференция. Если свет белый, то интерференционные полосы окрашены. Поэтому явление получило название *цветов тонких пленок.* Оно легко наблюдается на стенках мыльных пузырей, на тонких пленках масла или нефти, плавающих на поверхности воды, на пленках, возникающих на поверхности металлов при закалке (цвета побежалости), и т.д.

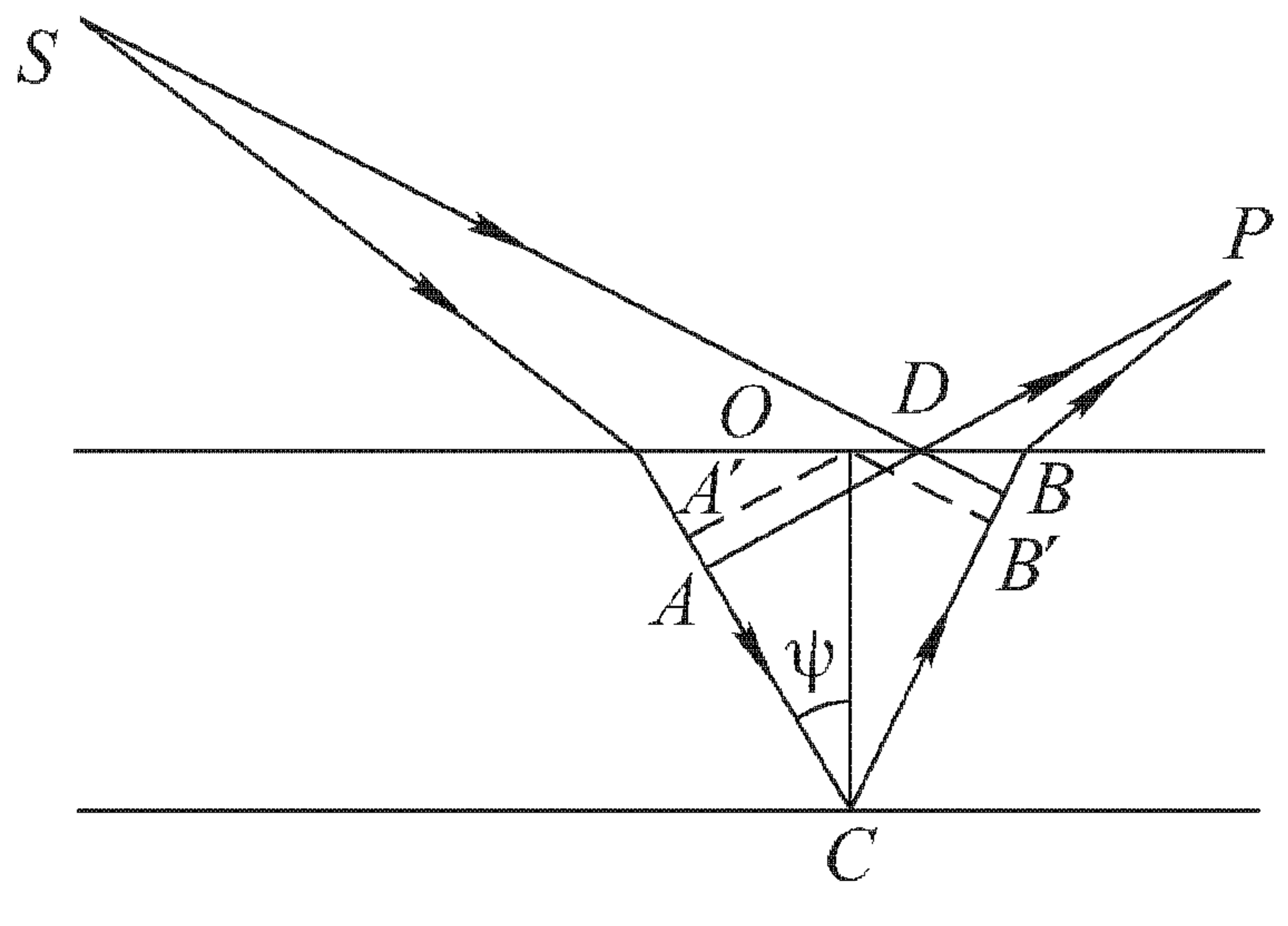


Рис. 133 *а*

2. Рассмотрим сначала плоскопараллельную пластинку толщины d с показателем преломления n, освещаемую точечным источником света S (рис. 133 *а*). При отражении от поверхностей пластинки возникает оптическая разность хода между соответствующими отраженными волнами. Вычислим ее в какой-либо произвольно выбранной точке P.

Пусть $SACBP$ и SDP — отраженные лучи, пересекающиеся в этой точке. Через точку D проведем плоскости DA и DB, перпендикулярные к соответствующим преломленным лучам в пластинке. Если пластинка тонкая, то для вычисляемой оптической разности хода можно написать: $\Delta = (SACBP) - (SDP)$. Через точку O, симметрично расположенную на верхней поверхности пластинки между точками A и B, проведем плоскости OA' и OB' (начерченные штриховыми линиями), параллельные плоскостям DA и DB. Тогда $\Delta = (A'CB') = 2(A'C) = 2nd\cos\psi$, где ψ — угол преломления.

Сюда необходимо, однако, ввести поправку. В пределе, когда толщина пластинки стремится к нулю, наша формула дает $\Delta = 0$. Поэтому в точке P должно было бы в этом случае происходить усиление колебаний. Но это невозможно, так как бесконечно тонкая пластинка вообще не может оказывать влияния на распространение света, так что никакого отражения возникнуть не может. Для этого волны, отраженные от передней и задней поверхностей пластинки, должны при интерференции гасить друг друга. Их фазы должны быть противоположны, т.е. оптическая разность хода Δ при $d \to 0$ должна стремиться к $\lambda/2$. Поэтому к прежнему выражению для Δ надо прибавить или отнять $\lambda/2$ (что совершенно безразлично). После этого получим [1]

$$\Delta = 2dn\cos\psi + \frac{\lambda}{2}. \qquad (33.1)$$

Введение такой поправки, как будет показано в § 65, обусловлено изменением фазы на π, которое претерпевает волна при отражении на одной из поверхностей пластинки. Экспериментально это явление подтверждается наблюдением интерференции белого света в установке Ллойда (см. конец § 30).

Все рассуждения и результаты остаются в силе и для случая, когда точка P лежит по другую сторону пластинки (рис. 133 *б*). В этом случае отраженные лучи расходящиеся, в точке P пересекаются не сами лучи, а их продолжения за пластинку. Для наблюдения интерференции в точке P надо отраженные лучи сделать сходящимися с помощью собирающей линзы или вогнутого зеркала. Тогда интерференцию следует наблюдать в точке P', оптически сопряженной с точкой P.

Формула (33.1) справедлива и для тонких пластинок переменной толщины. Только в этом случае под d надо понимать толщину пластинки в том ее месте, где происходит отражение лучей, пересекающихся в точке P. Но в точку P от точечного источника света могут попадать, вообще говоря, только два луча. Отсюда следует, что в случае точечного монохроматического источника света каждая точка пространства характеризуется вполне определенной разностью хода приходящих в нее отраженных лучей. Поэтому устойчивая интерференция должна наблюдаться *в каждой точке пространства.* Про соответствующие полосы интерференции говорят, что они *не локализованы* (или локализованы всюду).

---

[1] Неудовлетворительность приведенной аргументации была отмечена при выводе формулы (2.2).

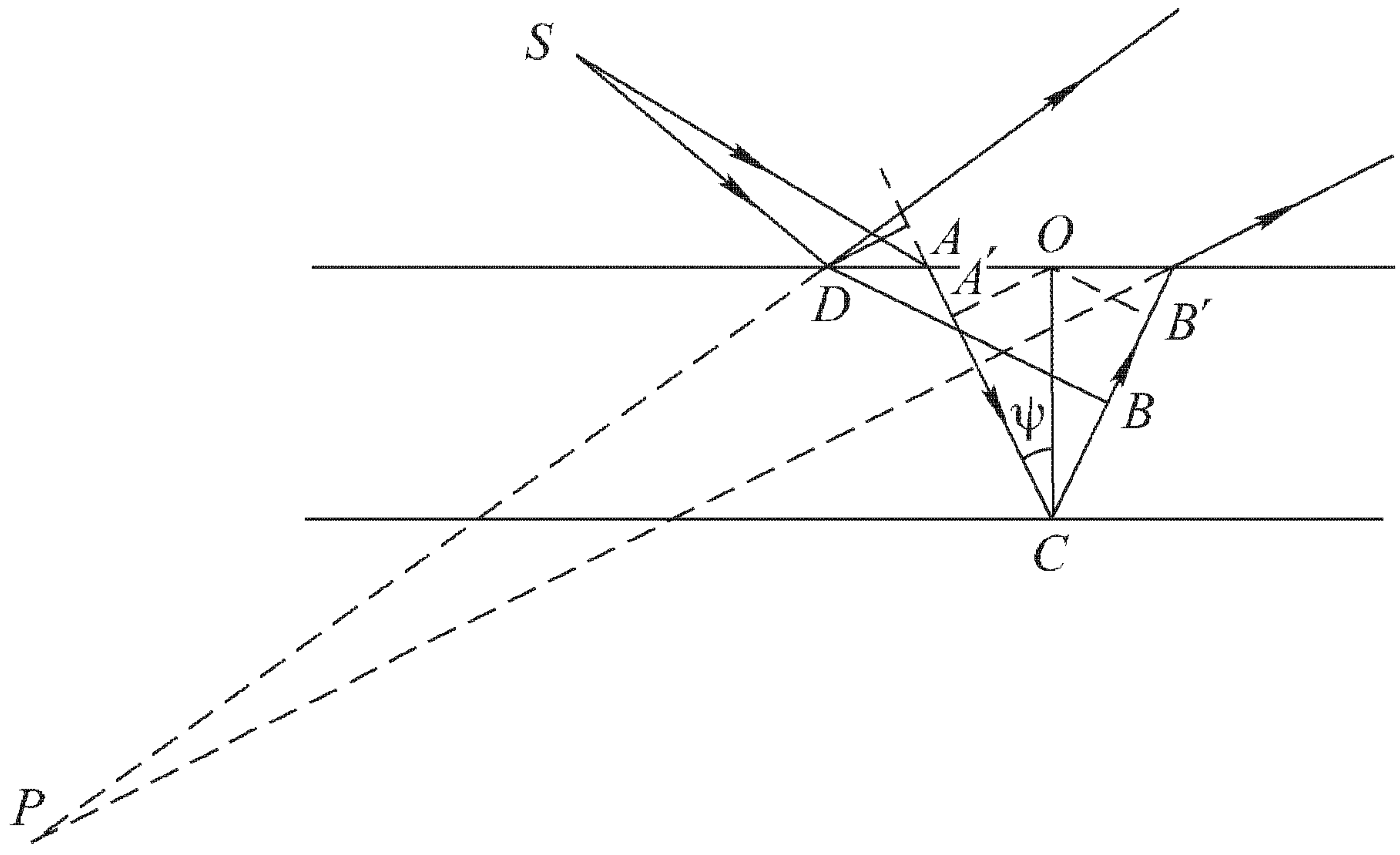


Рис. 133*б*

3. Положение существенно изменится, если источник света протяженный. Тогда в точку P будут попадать, вообще говоря, два отраженных луча от каждой точки источника. Место отражения на пластинке, углы падения и отражения, а следовательно, и разность хода для каждой пары лучей будут разными. Поэтому никакой устойчивой интерференции в точке P наблюдаться не будет.

Допустим, однако, что точка P находится близко от пластинки, например совпадает с точкой B. Тогда при малой толщине пластинки отражение будет происходить практически в одном и том же месте, т.е. при одной и той же толщине d, независимо от того, из какой точки источника исходят лучи. Если при этом лучи падают на пластинку почти нормально (так что $\cos\psi$ изменяется мало), то оптическая разность хода Δ будет зависеть только от толщины пластинки d в точке падения лучей. Линии на поверхности пластинки, где толщина d постоянна, будут также линиями постоянной разности фаз. На поверхности пластинки, если только степень монохроматичности света достаточна, появятся интерференционные полосы, каждая из которых характеризуется условием $d = \text{const}$. Они называются *полосами* или *линиями равной толщины.* Такие полосы как бы нарисованы на самой пластинке. Про них говорят, что они *локализованы на пластинке.* Конечно, интерференционные полосы должны наблюдаться не только на пластинке, но и с обеих сторон вблизи от нее. Однако максимальной видностью обладают интерференционные полосы на самой пластинке.

Для наблюдения таких полос удобно воспользоваться собирательной линзой, с помощью которой можно получить изображение пластинки на экране. Так как линза не вносит дополнительной разности хода, то при этом на экране получается изображение и интерференционных полос. Линза как бы переносит место локализации интерференционной картины с поверхности пластинки на экран. При визуальном наблюдении полос равной толщины глаз надо аккомодировать на пластинку. Роль линзы выполняет хрусталик, а экрана — сетчатая оболочка глаза. Оптический прибор или глаз выполняет также и другую полезную функцию. Диафрагма прибора или зрачок глаза вырезают из отраженных лучей узкие пучки, в пределах которых угол ψ меняется незначительно. Тем самым создаются условия, благоприятные для получения полос равной толщины.

При наблюдении в белом свете полосы интерференции окрашены, так как разность фаз $\dfrac{4\pi\, dn \cos\psi}{\lambda} + \pi$ зависит от λ. В белом свете возможна интерференция только *низкого порядка.* Поэтому пленка должна быть тонкой — ее толщина не должна превышать примерно одной-двух длин волн. Когда толщина пленки составляет доли длины волны, то разность фаз приближается к π, и пленка становится темной.

Если обе поверхности пленки плоские, то интерференционные полосы прямолинейны и параллельны линии пересечения соответствующих плоскостей. Такие полосы наблюдаются, например, в клине, т.е. тонкой воздушной прослойке между плоскопараллельными стеклянными пластинками, когда с одного края между ними проложен, например, тонкий лист бумаги. Но если поверхности сложенных стеклянных пластинок неровные, то полосы равной толщины принимают неправильную, причудливую форму. На этом основан чувствительный интерференционный *метод контроля* поверхности на плоскопараллельность. В этом методе испытуемая поверхность прижимается к плоской и наблюдаются полосы интерференции в образовавшейся прослойке. Метод применяется и для контроля сферических или параболических поверхностей при шлифовке оптических зеркал и линз.

4. Интерференция в тонких пленках рассматривалась выше как *двухлучевая интерференция.* Мы ограничились интерференцией только двух волн, одна из которых получилась при однократном отражении от верхней, а другая — от нижней поверхностей пленки. Многократными отражениями мы пренебрегли. Это можно делать, когда коэффициент отражения невелик. Если же коэффициент отражения близок к единице, то так поступать нельзя. Допустим, например, что при каждом отражении отражается 5 % падающего света, а остальные 95 % проходят. Примем интенсивность падающего света за единицу. Тогда, как нетрудно подсчитать, интенсивности отраженных и прошедших лучей (рис. 134) представятся следующей таблицей:

Лучи	1	2	3	4	1′	2′
Интенсивности	1	0,05	0,0451	0,00011	0,9025	0,00226

Из таблицы видно, что интенсивности отраженных лучей *2* и *3* почти одинаковы, а интенсивность луча *4* более чем в 100 раз меньше. Поэтому луч *4* и все отражения высших порядков можно не принимать во внимание. Из прошедших лучей интенсивность луча *2′* примерно в 400 раз меньше интенсивности луча *1′*. Поэтому интерференционные полосы в проходящем свете получаются *на светлом фоне* и по этой причине очень мало контрастны.

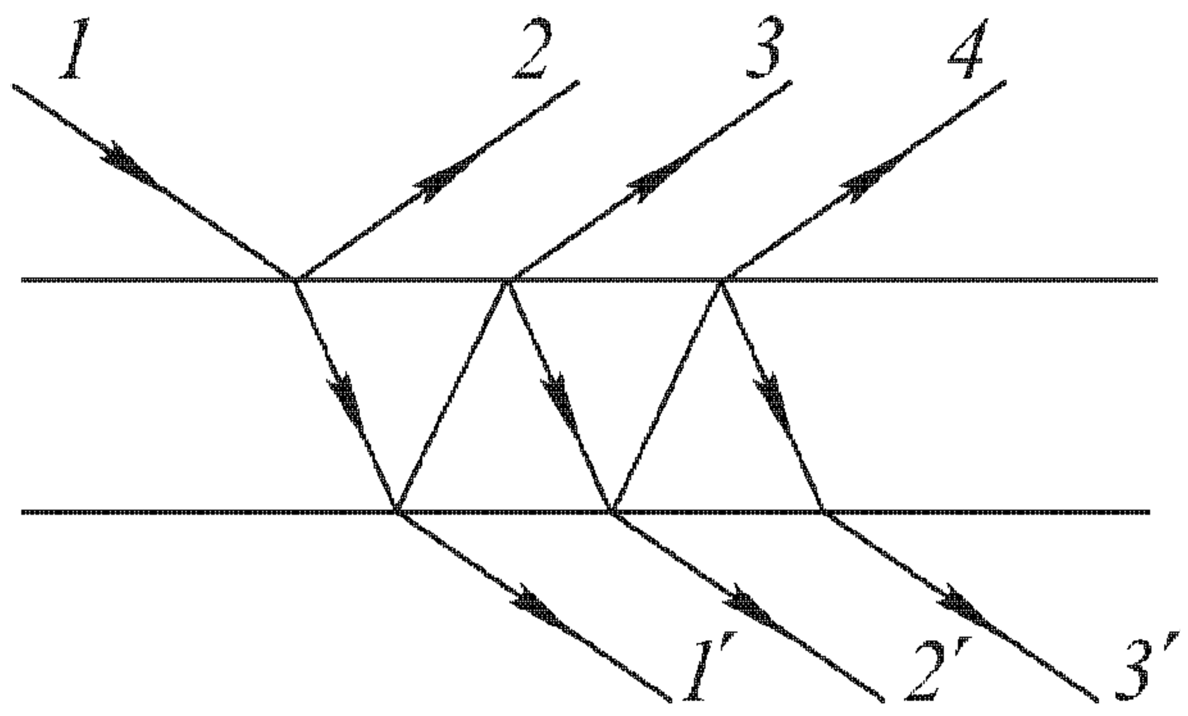


Рис. 134

5. Допустим теперь, что пластинка толстая и строго *плоскопараллельная*. Пусть она освещается параллельным пучком света. Формально это соответствует случаю точечного бесконечно удаленного источника S (см. рисунки 133 *а* и 133 *б*). Отраженные лучи будут также параллельными, т.е. точка наблюдения P удалится в бесконечность. При постоянной толщине пластинки d оптическая разность хода между отраженными лучами $2dn\cos\psi + \lambda/2$ зависит только от угла наклона падающих лучей. Угол ψ может принимать всевозможные значения, если источник света протяженный и имеет конечные угловые размеры. Практически это можно осуществить, поместив протяженный источник света в фокальной плоскости линзы, которая как бы удаляет источник света в бесконечность. Интерференционную картину следует наблюдать на бесконечно (т.е. достаточно) удаленном экране или в фокальной плоскости линзы, поставленной на пути отраженных лучей. Каждая интерференционная полоса на бесконечно удаленном экране характеризуется постоянством косинуса угла ψ. Поэтому интерференционные полосы при описанном способе наблюдения называют *полосами или линиями равного наклона*[1]. *Они локализованы в бесконечности.*

Линза переносит область локализации из бесконечности в фокальную плоскость, оптически сопряженную с бесконечно удаленной плоскостью. Роль линзы может играть глаз, аккомодированный на бесконечность. В этом случае полосы равного наклона локализуются на *сетчатке глаза*. При наблюдении в зрительную трубу последняя должа быть установлена на бесконечность.

Для толстых пластинок оптическая разность хода Δ велика, т.е. содержит тысячи и десятки тысяч длин волн. Интерференционные полосы будут высокого порядка. Для их получения требуется высокая степень монохроматичности $\lambda/\delta\lambda$ падающего света.

6. Ньютон наблюдал интерференционные полосы равной толщины в воздушной прослойке между плоской поверхностью стеклянной пластинки и плоско-выпуклой линзой, прижатой к пластинке выпуклой стороной. Интерференционные полосы имели форму концентрических колец. Они получили название *колец Ньютона*. При постановке опыта радиус кривизны выпуклой поверхности линзы надо брать не менее приблизительно одного метра. Кольца удобно наблюдать в длиннофокусный микроскоп, сфокусированный на воздушную прослойку. Более контрастные кольца наблюдаются, конечно, в отраженном, а не в проходящем свете. Чтобы наблюдалось много колец, надо пользоваться светом сравнительно высокой монохроматичности. Подходящим может быть желтый свет натровой горелки или свет ртутной лампы. Вид колец показан на рис. 135.

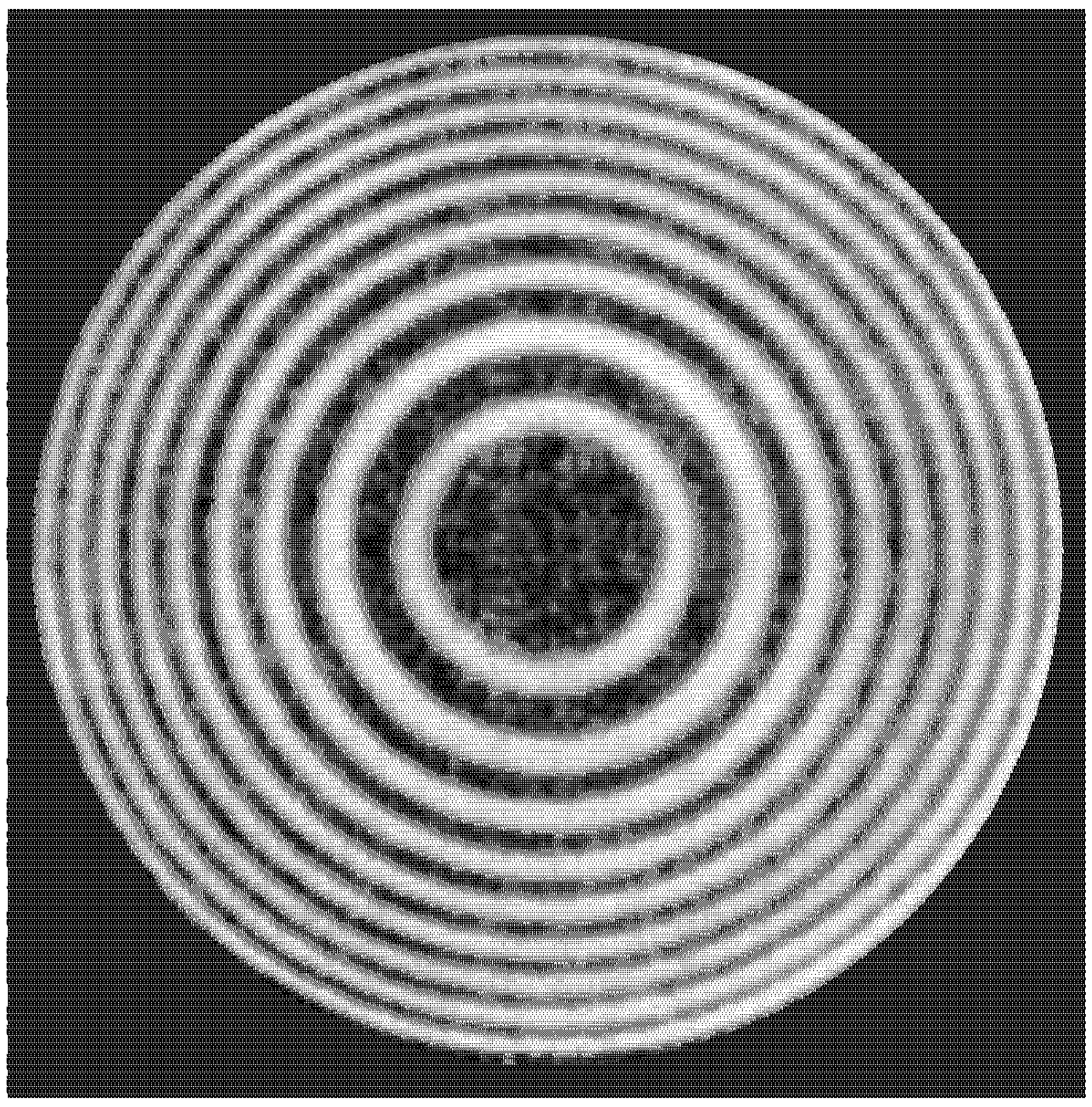


Рис. 135

Для вычисления радиусов колец дополним выпуклую поверхность линзы до полной сферы (рис. 136). Если BD — диаметр, то по известной геометрической теореме $AB \cdot AD = AO^2 = x^2$. Ввиду малости кривизны сферы и угла OCA, отрезок AB можно принять за толщину d воздушной прослойки в точке A. Пренебрегая еще различием между AD и диаметром сферы $2R$, получим

---

[1] Этот общепринятый термин не совсем точно передает существо дела. Лучам одинакового наклона соответствует не линия, а точка — фокус линзы, в котором они собираются.

$$d = \frac{x^2}{2R}, \qquad \Delta = 2d + \frac{\lambda}{2} = \frac{x^2}{R} + \frac{\lambda}{2}.$$

Светлые кольца получатся при $\Delta = m\lambda$, где m — целое число. Из этого условия находим радиус x_m m-го светлого кольца:

$$x_m = \sqrt{(m - {}^1\!/_2)\lambda R} = \sqrt{\lambda R/2}\sqrt{2m-1}. \qquad (33.2)$$

Аналогично для радиуса m-го темного кольца

$$x_m = \sqrt{m\lambda R} = \sqrt{\lambda R/2}\sqrt{2m}. \qquad (33.3)$$

Таким образом, радиусы последовательных светлых колец пропорциональны квадратным корням из нечетных чисел 1, 3, 5, ..., а радиусы темных колец — квадратным корням из последовательных четных чисел 0, 2, 4, ... Эти закономерности экспериментально были установлены Ньютоном. Как указывалось в § 3 (п. 6), в интерференции Ньютон видел проявление *периодичности* световых процессов. Из наблюдения интерференционных колец Ньютон даже довольно точно вычислил величину, являющуюся количественной мерой указанной периодичности. В переводе на язык волновой теории вычисленная Ньютоном величина есть половина длины световой волны.

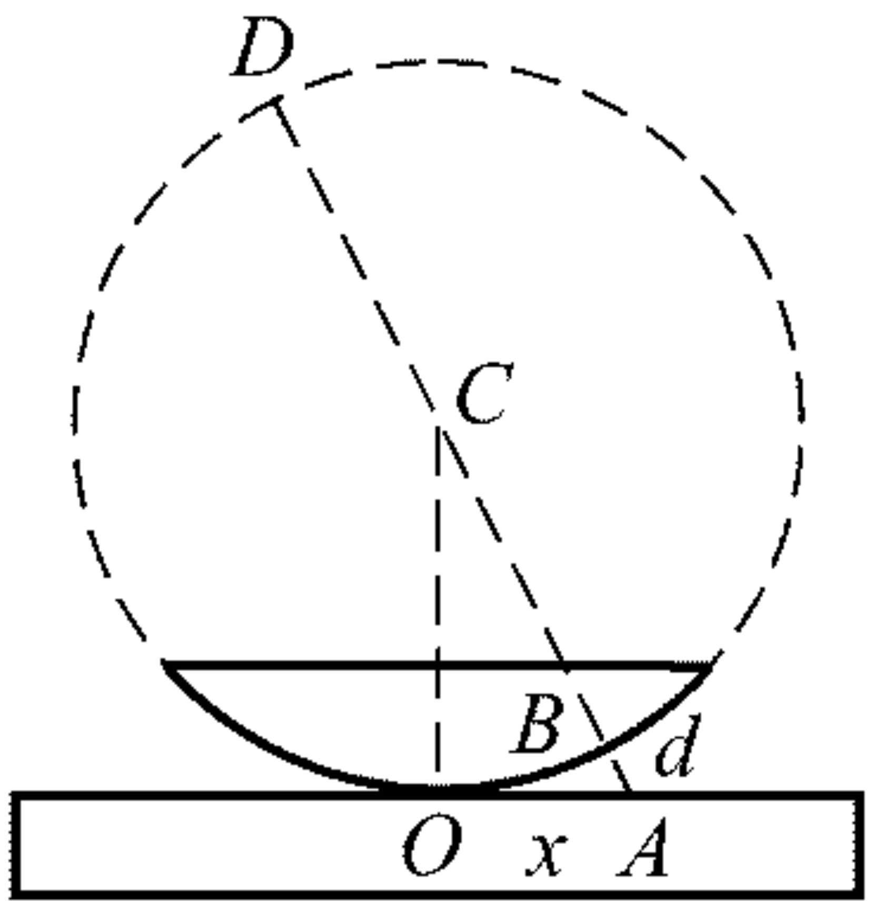


Рис. 136

Центр колец в отраженном свете темный, а в проходящем светлый. Это доказывает, что при отражении на одной из границ воздушной прослойки фаза отраженной волны меняется на π. Как будет показано в § 65, электрический вектор меняет фазу на π при отражении от среды с бо́льшим показателем преломления. В противоположном случае изменения фазы не наблюдается. Это подтверждается следующим опытом Юнга. Он взял пластинку из флинта (n = 1,7), прижал к ней линзу из крона (n = 1,5), а прослойку между ними заполнил сассафрасовым маслом, показатель преломления которого имел промежуточное значение. Тогда фаза волны менялась на π при отражении как от верхней, так и от нижней поверхностей масляной прослойки. Благодаря этому центр ньютоновых колец получался светлым в отраженном и темным в проходящем свете.

7. Если в установке Ньютона линзу перемещать вверх параллельно самой себе, то из-за увеличения толщины воздушной прослойки каждая окружность постоянной (фиксированной) разности хода будет стягиваться к центру картины. Интерференционные кольца также станут стягиваться к центру, так как вдоль каждого кольца оптическая разность хода остается постоянной. Достигнув центра, каждое кольцо превратится в кружок, исчезающий при дальнейшем перемещении линзы. Таким образом, центр картины будет становиться попеременно то светлым, то темным. Одновременно на периферии поля зрения будут зарождаться и перемещаться к центру новые интерференционные кольца, пока каждое из них не исчезнет в центре картины. Можно сказать, что при перемещении линзы непрерывно вверх пропадают кольца самых низких порядков интерференции и зарождаются кольца более высоких порядков. Таким путем удается наблюдать интерференцию все более и более высокого порядка. Понятно, что для этого требуется свет высокой степени монохроматичности.

Физо, впервые осуществивший подобный опыт, пользовался желтым светом натровой горелки (в опыте Физо интерференционные полосы получались между двумя плоскопараллельными пластинками). Физо заметил, что при увеличении числа полос N, прошедших в поле зрения, ухудшалась видимость полос. При $N = 490$ видимость достигала минимума, затем она опять улучшалась; при $N = 980$ полосы приобретали прежнюю отчетливость. Ближайший минимум наступал при $N = 1470$, следующий максимум при $N = 1960$ и т.д. Физо сумел заметить 52 таких максимума. На этом основании он пришел к заключению, что желтая линия натрия — двойная, т.е. состоит из двух близко расположенных спектральных линий с длинами волн λ и $\lambda' > \lambda$.

Легко найти расстояние между этими линиями $\lambda' - \lambda$. Как было показано в § 30, первое исчезновение видимости интерференционных полос происходит при условии $\lambda' - \lambda = \lambda/2N$. По наблюдениям Физо $N = 490$. Средняя длина волны желтой спектральной линии натрия $\approx 589{,}3$ нм. Поэтому $\lambda' - \lambda \approx 0{,}6$ нм.

Этот пример показывает, как интерференция высокого порядка может быть использована для разрешения сложных спектральных линий, состоящих из двух близко расположенных линий.

§ 34. Интерферометр Жамена

1. *Интерферометрами* называют оптические измерительные приборы, основанные на интерференции света. Они позволяют с высокой степенью точности измерять линейные и угловые расстояния, малые разности показателей преломления, исследовать структуру спектральных линий и т.п. Принцип действия интерферометров в основном один и тот же. Но в зависимости от предназначения им придают различные конструкции.

2. *Интерферометр Жамена* (1818–1886) состоит из двух толстых (толщина не менее 20 мм) плоскопараллельных пластин P_1

и P_2, изготовленных из весьма однородного стекла и смонтированных на массивной плите (рис. 137). Для исследований в далекой ультрафиолетовой области спектра пластины изготовляются из кварца или флюорита. Чтобы обе пластины имели строго одинаковую толщину, они вырезаются из одной и той же пластины. Пластины устанавливаются приблизительно под углом 45° к линии, соединяющей их центры. Угол между пластинами можно менять с помощью специальных установочных винтов, поворачивающих пластины вокруг вертикальной и горизонтальной осей. Чем меньше этот угол, тем шире интерференционные полосы. Задние отражающие поверхности пластин посеребрены. Луч света SA, отразившись от передней и задней поверхностей пластины P_1, разделяется на два параллельных луча AB и $C'D'$, между которыми возникает разность хода. Каждый из этих лучей в свою очередь раздваивается при отражении от пластины P_2. Между средними лучами DE и $D'E'$ возникает разность хода

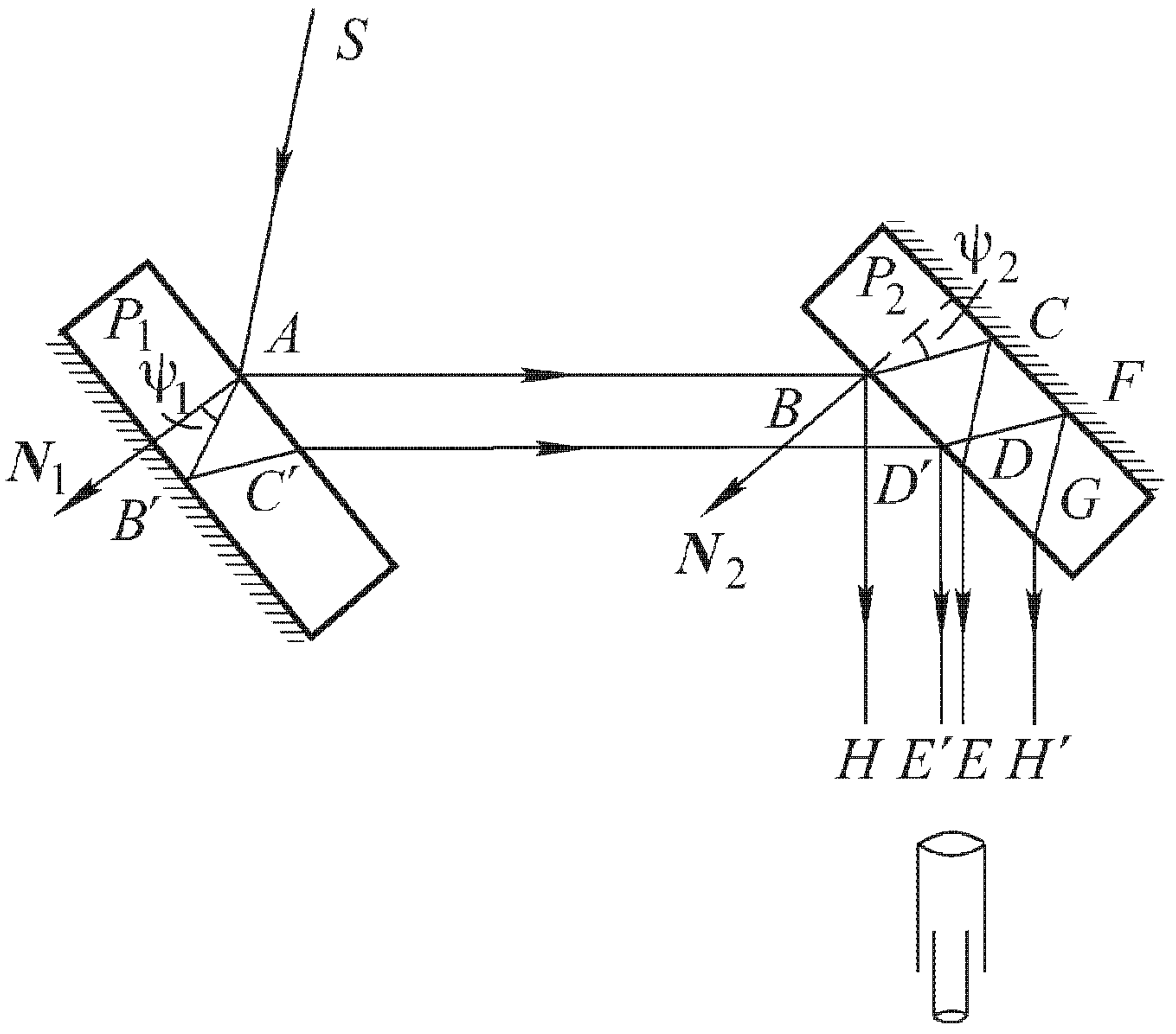


Рис. 137

$$\Delta = 2dn(\cos\psi_2 - \cos\psi_1), \tag{34.1}$$

где d — толщина пластин, ψ_1 и ψ_2 — углы преломления в них, n — показатель преломления пластин относительно окружающего воздуха. Таким образом, на величину Δ пластины влияют противоположно. Потери фаз при отражении в формулу (34.1) не входят, так как они одинаковы для обоих интерферирующих лучей и компенсируют друг друга. Если $\psi_1 = \psi_2$, то $\Delta = 0$. Если же угол между пластинами мал, то и разность хода невелика. Это позволяет наблюдать интерференцию низкого порядка, используя белый свет. Лучи DE и $D'E'$ будут интерферировать так же, как и соответствующие лучи в случае только одной, но тонкой плоскопараллельной пластинки.

Назначение двух толстых пластин состоит в том, чтобы далеко развести лучи AB и $C'D'$, что позволяет ставить на пути этих лучей трубки с исследуемыми газами. Разность хода между средними и крайними лучами, а также между самими крайними лучами очень велика, и интерференцию в белом свете от этих лучей получить невозможно. Источником света обычно служит освещаемая вертикальная щель шириной около одного или нескольких миллиметров, прикрываемая матовым стеклом. Невооруженным глазом видны три изображения щели. Самое яркое (справа) дают лучи, дважды отразившиеся от задних посеребренных поверхностей пластин. Среднее, промежуточной яркости, образуется лучами DE и $D'E'$. Наименее яркое (слева) соответствует лучам, дважды отразившимся от передних поверхностей пластин. Если фиксировать глаз на среднем изображении и вращением установочных винтов менять разность хода, то в белом свете можно получить цветные интерференционные полосы низкого порядка. Центральная ахроматическая полоса, т.е. полоса, которой соответствует разность хода $\Delta = 0$, получается белой. Наблюдая эту полосу, можно легко следить за смещением интерференционной картины.

Так как разность хода Δ зависит от направления падающих лучей, то в интерферометре Жамена получаются полосы равного наклона. Для их наблюдения удобно воспользоваться зрительной трубой небольшого увеличения, но большого поля зрения, установленной на бесконечность. Применение трубы не только увеличивает ширину полос, но и обладает еще тем преимуществом, что в трубу не попадают крайние лучи GH' и BH, создающие светлый фон и сильно уменьшающие контрастность интерференционных полос. В лекционных демонстрациях интерференционные полосы можно получать без дополнительных оптических приспособлений, направив лучи DE и $D'E'$ непосредственно на белый экран. Крайние лучи при этом должны быть экранированы.

3. Интерферометр Жамена предназначен для измерения небольших изменений показателей преломления. Поэтому его называют также *интерференционным рефрактометром.* Для уяснения принципа действия такого рефрактометра вообразим, что на пути одного из интерферирующих лучей AB или $C'D'$ (см. рис. 137) помещен плоскопараллельный слой какого-либо вещества толщины l с показателем преломления n_2. Тогда разность хода между интерферирующими лучами изменится на величину $(n_2 - n_1)l$, где n_1 — показатель преломления окружающего воздуха. В результате интерференционная картина сместится на m полос, причем $m = (n_2 - n_1)l/\lambda$. Число m (вообще говоря, дробное) можно найти, наблюдая интерференционные полосы в белом свете до и после внесения исследуемого вещества. Опыт показывает, что смещение на 1/10 полосы ($m = 1/10$) наблюдается вполне уверенно и без труда.

Разность показателей преломления можно вычислить по формуле $n_2 - n_1 = m\lambda/l$. Если толщина слоя вещества $l = 10$ см, а $\lambda = 500$ нм, то при $m = 1/10$ получаем $n_2 - n_1 = 5\cdot 10^{-7} \sim 10^{-6}$. В специальных установках измерялись гораздо меньшие изменения показателя преломления. Интерферометр Жамена позволяет, таким образом, измерять ничтожные изменения показателя преломления, например, при изменении температуры газа или при добавлении к нему посторонних примесей. Измерение самих абсолютных значений показателя преломления этим прибором затруднительно. Обычно его используют для измерения разности показателей преломления исследуемого и хорошо изученного газа, например воздуха.

При практических измерениях на пути лучей AB и $C'D'$ ставятся по возможности одинаковые трубки. Сначала наблюдаются интерференционные полосы, когда обе трубки наполнены воздухом. Затем измеряется смещение полос интерференции, когда одна из трубок наполняется исследуемым газом. Для компенсации значительной разности хода между интерферирующими лучами служит компенсатор. Он состоит из двух совершенно одинаковых стеклянных пластинок, насаженных на общую ось, причем угол наклона между пластинками можно плавно менять. Одна из пластинок ставится на пути одного, вторая — на пути другого лучей. Вращая компенсатор при неизменном угле наклона между пластинками, можно привести в поле зрения центральную ахроматическую полосу, а затем по углу поворота определить число полос, прошедших через поле зрения.

4. Не так просто разобраться в вопросе о форме интерференционных полос, наблюдаемых в интерферометре Жамена. Для упрощения исследования этого вопроса введем несущественное предположение, что $n = 1$. Такому предположению соответствует замена каждой пластины парой бесконечно тонких параллельных отражающих плоскостей (из которых одна посеребрена). Тогда вместо (34.1) получится более простое выражение:

$$\Delta = 2d(\cos\psi_2 - \cos\psi_1).$$

Перейдем к векторной форме записи, обозначив через $\mathbf{N}_1$ и $\mathbf{N}_2$ единичные нормали к отражающим плоскостям, как это показано на рис. 137. Обозначим через $\mathbf{s}_1$, $\mathbf{s}_1'$, $\mathbf{s}$ единичные векторы лучей SA, AB и BH соответственно (те же векторы определяют направления и всех параллельных им лучей). Тогда

$$\cos\psi_1 = (\mathbf{N}_1\mathbf{s}_1') = -(\mathbf{N}_1\mathbf{s}_1'), \qquad \cos\psi_2 = (\mathbf{N}_2\mathbf{s}).$$

По закону отражения $\mathbf{s}_1' = \mathbf{s} - 2(\mathbf{N}_2\mathbf{s})\mathbf{N}_2$, а потому

$$\cos\psi_1 = -(\mathbf{N}_1\mathbf{s}) + 2(\mathbf{N}_2\mathbf{s})(\mathbf{N}_1\mathbf{N}_2).$$

Следовательно,

$$\Delta = 2d(\mathbf{N}_1 + \mathbf{N}_2)\mathbf{s} - 4d(\mathbf{N}_2\mathbf{s})(\mathbf{N}_1\mathbf{N}_2).$$

Введем новые векторы

$$\mathbf{N} = \frac{1}{2}(\mathbf{N}_1 + \mathbf{N}_2), \quad \boldsymbol{\nu} = \frac{1}{2}(\mathbf{N}_2 - \mathbf{N}_1).$$

Из них первый направлен по биссектрисе угла между нормалями $\mathbf{N}_1$ и $\mathbf{N}_2$, а второй перпендикулярен к этой биссектрисе. Очевидно,

$$\mathbf{N}_1 = \mathbf{N} - \boldsymbol{\nu}, \quad \mathbf{N}_2 = \mathbf{N} + \boldsymbol{\nu}.$$

После подстановки этих выражений в формулу для Δ получим

$$\Delta = -4d(1 - \boldsymbol{\nu}^2)(\boldsymbol{\nu}\mathbf{s}) + 4d\boldsymbol{\nu}^2. \tag{34.2}$$

При работе с прибором угол между пластинами, а следовательно, и длина вектора $\boldsymbol{\nu}$ всегда малы. Поэтому квадратичными членами, содержащими $\boldsymbol{\nu}^2$, в формуле (34.2) обычно можно пренебречь.

5. Для светлых полос $\Delta = m\lambda$, для темных $\Delta = (m+{}^1\!/_2)\lambda$, где m — целое число (порядок интерференции). Если для наблюдения используется труба, то она сводит параллельные лучи в фокальной плоскости объектива, где и получается первичная интерференционная картина. Поле зрения трубы ограничено, так что используется лишь небольшой участок фокальной плоскости. При повороте трубы вокруг оптического центра объектива поворачивается и фокальная плоскость вместе с выделенным участком. Если последний мал, то без заметной ошибки можно принять все такие участки за элементы неподвижной сферы, центр которой совпадает с оптическим центром объектива. Назовем для краткости такую вспомогательную сферу *фокальной сферой объектива*. (Если объектив не является тонкой линзой, то все это можно повторить, заменив только оптический центр второй узловой точкой объектива.) Можно сказать, что первичная интерференционная картина как бы создается на неподвижной фокальной сфере, а наблюдатель рассматривает эту картину через окуляр трубы. Поворачивая трубу, он видит на фокальной сфере различные участки интерференционной картины. Разумеется, вид картины совершенно не зависит от того, вокруг какой точки вращалась труба, он определяется только *направлением* последней.

Таким образом, вопрос о форме видимых полос интерференции сводится к вопросу о форме их на вспомогательной фокальной сфере. Практически, однако, вращение трубы всегда производится вокруг вертикальной оси. Поэтому наблюдению доступен только узкий экваториальный поясок на фокальной сфере. Его приближенно можно считать поверхностью цилиндра с вертикальной осью, проходящей через оптический центр объектива. В дальнейшем ограничимся рассмотрением интерференционных полос только на таком «цилиндрическом фокальном пояске».

Фиксировав положение пластин интерферометра, отложим векторы $\mathbf{N}$ и $\boldsymbol{\nu}$, а также единичный вектор луча $\mathbf{s}$ из общего начала O, помещенного в оптическом центре объектива. Направления этих векторов будут играть роль координатных осей, перпендикулярную к ним прямую примем за ось Z. Различные единичные векторы лучей $\mathbf{s}$ будем откладывать из того же начала O. Как видно из (34.2), на поверхностях равных фаз $(\boldsymbol{\nu}\mathbf{s}) = \text{const}$, т.е. эти поверхности представляют собой параллельные плоскости, перпендикулярные к вектору $\boldsymbol{\nu}$. Среди этих плоскостей находятся и равноотстоящие плоскости, на которых разность хода Δ содержит целое или полуцелое число длин волн. Они пересекают фокальную сферу вдоль некоторых кривых, которые и представляют собой интерференционные полосы, наблюдаемые в трубу. Ввиду малости поля зрения, при каждом положении трубы наблюдается лишь незначительный участок фокальной сферы, на котором интерференционные полосы практически неотличимы от прямолинейных.

В интерферометре Жамена первая пластина обычно неподвижна и установлена вертикально. Вторую пластину можно немного поворачивать вокруг вертикальной и горизонтальной осей, параллельных отражающим плоскостям. Если пластины строго параллельны друг другу, то $\boldsymbol{\nu} = 0$ и $\Delta = 0$. Следовательно, никаких интерференционных полос нет, все поле зрения освещено равномерно. В этом положении нормали $\mathbf{N}_1$ и $\mathbf{N}_2$, а следовательно, и вектор $\mathbf{N}$ лежат в горизонтальной плоскости.

Повернем теперь вторую пластину на малый угол вокруг вертикальной оси. Тогда вектор $\boldsymbol{\nu}$ расположится в той же горизонтальной плоскости (рис. 138 *а*). Перпендикулярные к нему плоскости равных фаз вертикальны. Одна из таких плоскостей $BCEF$ изображена на рис. 138 *а*. Она пересекает экваториальный цилиндрический поясок вдоль вертикального отрезка EF. Следовательно, интерференционные полосы, лежащие на этом пояске, будут вертикальны. На схематическом рис. 138 *а* они представлены жирными отрезками. Ширина интерференцион-

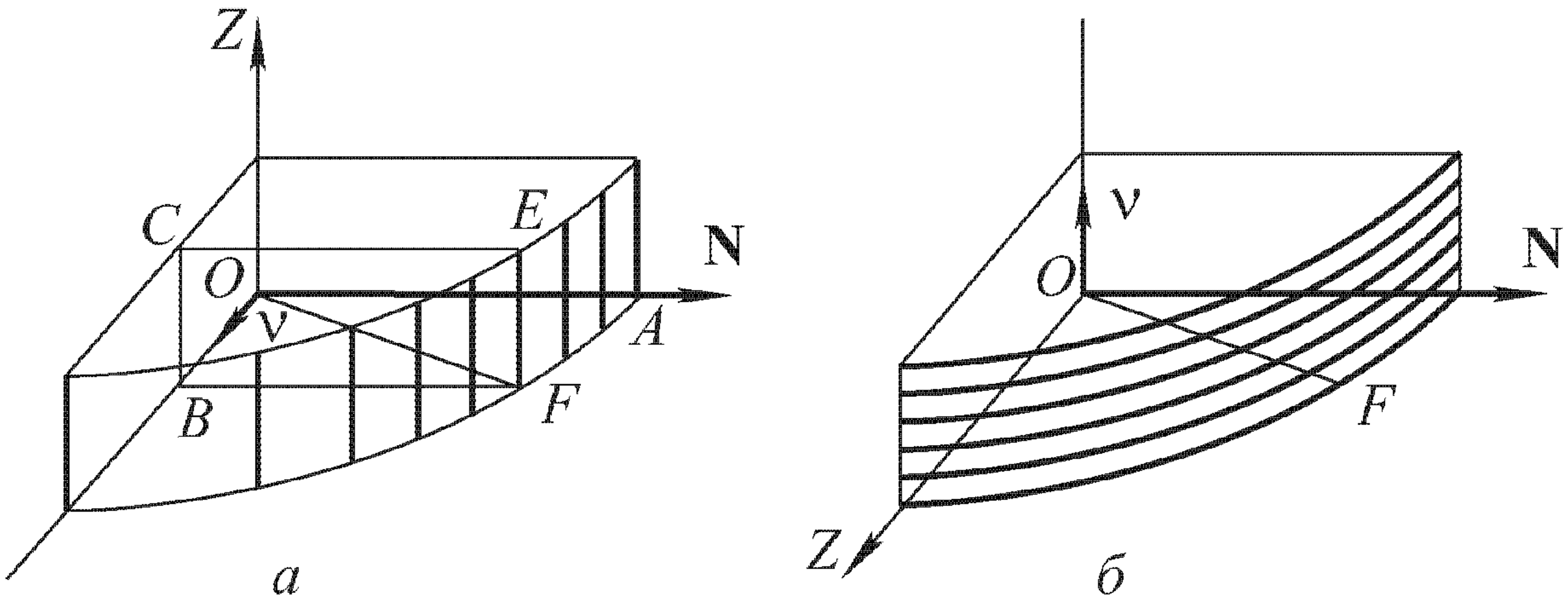


Рис. 138

ных полос минимальна при наблюдении вдоль $\mathbf{N}$ и максимальна при наблюдении вдоль $\mathbf{v}$. Однако фактически ось трубы OF бывает направлена под углом $\approx 45°$ к нормали $\mathbf{N}$. Тогда вертикальные интерференционные полосы могут наблюдаться только в монохроматическом (квазимонохроматическом) свете, так как им соответствуют высокие порядки интерференции. Действительно, если наблюдение ведется вдоль $\mathbf{N}$, то $(\mathbf{sv}) = 0$, и формула (34.2) дает $\Delta = 4d\mathbf{v}^2$. Ввиду малости $\mathbf{v}$ порядок интерференции получается низким. Нулевая полоса практически проходит через точку A.

Когда труба направлена вдоль $\mathbf{N}$ или близко к этому направлению, вертикальные интерференционные полосы можно увидеть и в белом свете. Но при удалении от этого направления скалярное произведение $(\mathbf{vs})$, а с ним и порядок интерференции возрастают. При повороте на угол $\approx 45°$ порядок интерференции становится столь высоким, что вертикальные интерференционные полосы в белом свете наблюдаться уже не могут. Исключение было бы только в случае очень малых $\mathbf{v}$, когда порядок интерференции при любых положениях трубы низкий. Но тогда интерференционная полоса была бы столь широка, что не уместилась бы в поле зрения трубы. Поэтому этот случай не представляет никакого интереса.

Совсем иное положение возникнет, если вторую пластину из исходного положения повернуть вокруг горизонтальной оси. Тогда вектор $\mathbf{v}$ встанет вертикально (рис. 138 *б*), а перпендцкулярные к нему плоскости равных фаз сделаются горизонтальными. Интерференционные полосы определятся, если пересечь такими плоскостями поверхность экваториального цилиндрического пояска. На рис. 138 *б* они изображены жирными дугами окружностей, плоскости которых горизонтальны. При этом все они будут низкого порядка, а потому могут наблюдаться в белом свете, так как наблюдение ведется практически перпендикулярно к $\mathbf{v}$, когда $(\mathbf{vs}) \approx 0$. В поле зрения трубы будут видны практически прямолинейные горизонтальные полосы. Легко сообразить, что при наклоне вектора $\mathbf{v}$ интерференционные полосы сделаются также наклонными к горизонту. Но в белом свете могут наблюдаться только интерференционные полосы малого наклона. Для получения полос с большим наклоном надо пользоваться монохроматическим (квазимонохроматическим) светом.

Все приведенные рассуждения можно повторить по той же схеме и не вводя предположения, что $n = 1$. Для избежания громоздкости вычислений надо только всюду ограничиться членами, линейными по $\mathbf{v}$, отбрасывая все члены более высоких степеней. При этом окончательные результаты не претерпят никаких существенных изменений.

6. Интерферометр Жамена обладает рядом недостатков, делающих его непригодным в некоторых особенно тонких оптических исследованиях. Изготовить пластины толще 5 см из вполне однородного стекла с точно параллельными плоскостями практически невозможно. Поэтому и интерферирующие лучи AB и $C'D'$ невозможно развести на расстояние больше $\sim$ 4 см. Но главный недостаток интерферометра Жамена состоит в том, что толстые пластины при освещении медленно прогреваются и медленно приходят в состояние теплового равновесия. Это приводит к тому, что интерференционные полосы часами медленно ползут в поле зрения трубы. Крайне затруднена работа в ультрафиолетовой области, требующая уникальных пластин из кварца или флуорита. Указанные недостатки устранены в интерферометре Д.С. Рождественского (1876–1940).

Интерферометр Рождественского принципиально не отличается от интерферометра Жамена. Каждая пластина в нем заменена комбинацией из двух параллельно установленных тонких пластинок M_1, P_1 и M_2, P_2 (рис. 139). Пластинки M_1 и M_2 посеребрены и выполняют роль зеркал. Пластинки P_1 и P_2, также посеребрены, полупрозрачны, они пропускают половину падающего света, а остальную половину отражают. Комбинация параллельных пластинок M_2, P_2 может поворачиваться на малый угол относительно другой комбинации также параллельных пластинок M_1, P_1. Ход лучей в интерферометре показан на схематическом рис. 139. Прибор действует так же, как и интерферометр Жамена. В нем наблюдаются интерференционные полосы равного наклона.

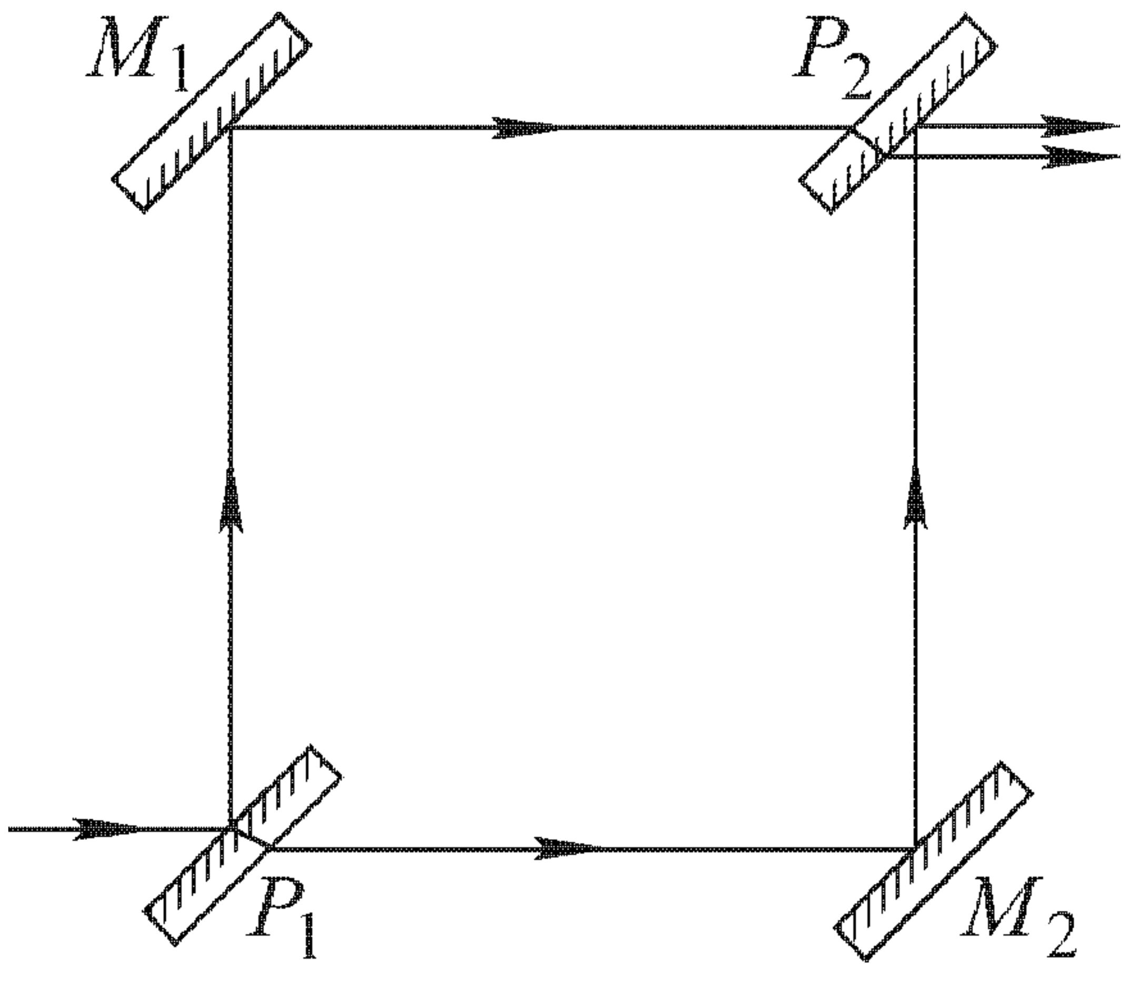


Рис. 139

Аналогично устроен построенный несколько раньше интерферометр Маха–Цендера. Отличие от интерферометра Рождественского состоит в том, что параллельно устанавливаются между собой зеркала M_1 и M_2, а также полупрозрачные пластинки P_1 и P_2. Одна пара пластин может поворачиваться на малый угол относительно другой пары.

§ 35. Интерферометр Майкельсона

С помощью *интерферометра Майкельсона* (1852–1931) впервые было проведено систематическое изучение тонкой структуры спектральных линий и выполнено первое прямое сравнение эталонного метра с длиной световой волны. Знаменитый опыт

Майкельсона–Морли, целью которого было обнаружение движения Земли относительно эфира (см. § 102), был выполнен также с помощью интерферометра Майкельсона. В настоящее время интерферометр Майкельсона устарел. Однако историческое значение указанных выше исследований, в особенности последнего, слишком велико. Поэтому необходимо хотя бы кратко остановиться на описании устройства и работы интерферометра Майкельсона.

Схема интерферометра приведена на рис. 140. Свет от протяженного источника S попадает на плоскопараллельную разделительную пластинку P_1, покрытую полупрозрачным тонким слоем серебра или алюминия. Эта пластинка частично пропускает, частично отражает свет, разделяя падающий пучок на два взаимно перпендикулярных пучка. Первый пучок, пройдя через пластинку P_1, отражается обратно зеркалом M_1, а затем частично отражается от пластинки P_1 в направлении AO. Второй пучок, отразившись от покрытой металлом поверхности пластинки P_1, направляется к зеркалу M_2, отражается от него, снова проходит через пластинку P_1 и далее идет в направлении AO, как и первый пучок.

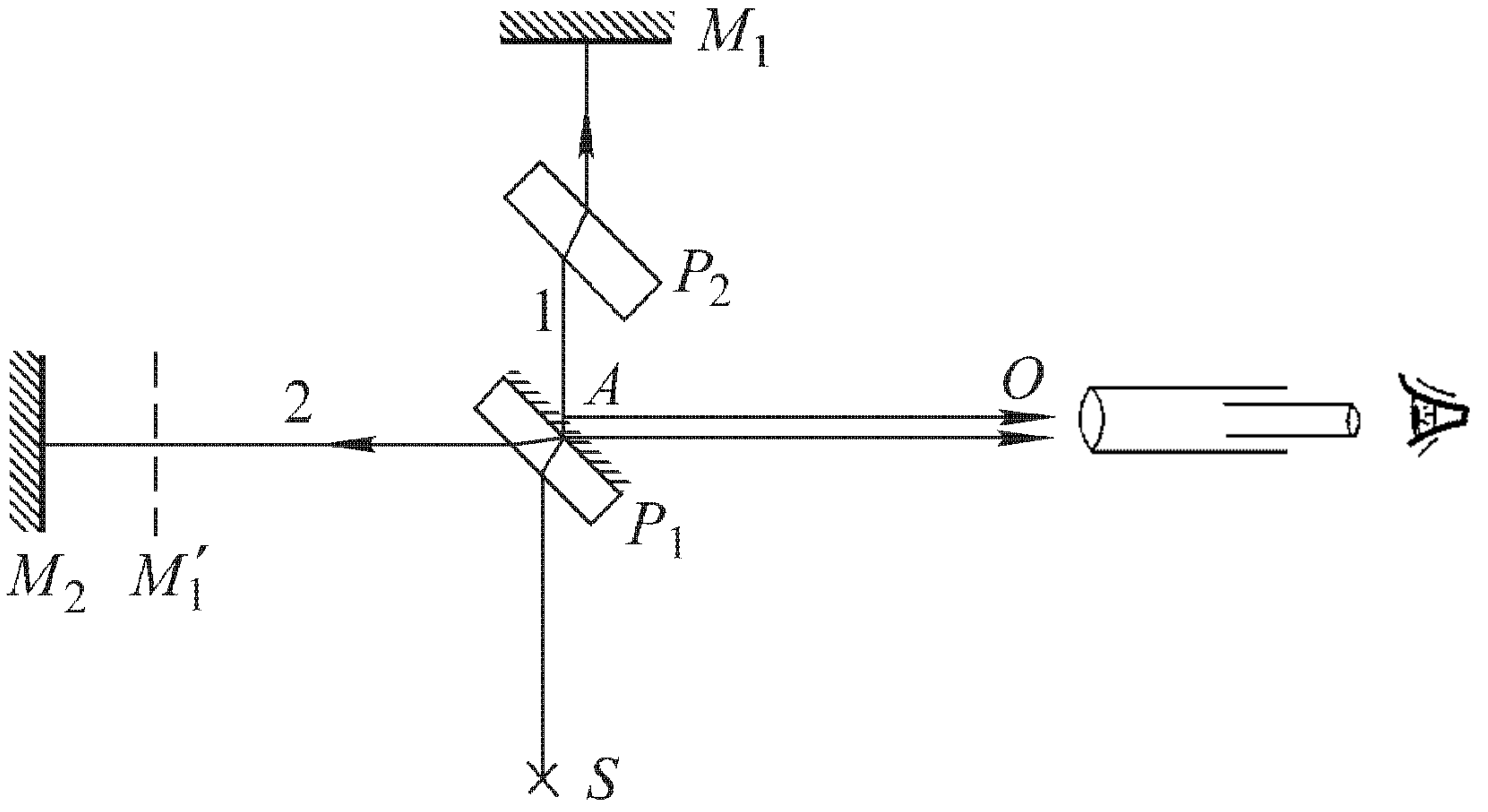


Рис. 140

Таким образом, от одного и того же источника получаются два пучка лучей одинаковой интенсивности, идущих в зрительную трубу, где и наблюдаются интерференционные полосы.

На пути первого пучка лучей ставится пластинка P_2, тождественная пластинке P_1. Она компенсирует разность хода между пучками, возникающую из-за того, что второй пучок пересекает разделительную пластинку P_1 три, а первый только один раз. Так как стекло обладает дисперсией, то без такой компенсации наблюдение интерференции в белом свете было бы невозможно.

Зеркало M_1 неподвижно, а M_2 с помощью микрометрического винта может перемещаться на специальных салазках строго параллельно самому себе. В больших интерферометрах такое перемещение должно происходить на десятки сантиметров. Поэтому к механической части прибора предъявляются исключительно высокие требования. Зеркала интерферометра снабжены также установочными винтами, позволяющими придавать им правильное положение.

Пусть M_1' — изображение поверхности зеркала M_1 в отражающей плоскости разделительной пластинки P_1. Тогда интерференция будет происходить так же, как и в воздушном слое между двумя отражающими плоскостями M_2 и M_1'. Разность хода между отраженными лучами $\Delta = 2d\cos\varphi$, где d — толщина слоя, а φ — угол падения. Если слой плоскопараллелен, то будут получаться интерференционные полосы равного наклона, локализованные в бесконечности. Их можно наблюдать глазом, аккомодированным на бесконечность, или в трубу, установленную также на бесконечность. Получатся интерференционные кольца с центром в точке схождения лучей, нормально отраженных от поверхностей M_2 и M_1'. Этому направлению соответствует максимальная разность хода $\Delta = 2d$. Поэтому максимальный порядок интерференции будет наблюдаться в центре картины. Отсюда следует, что при увеличении толщины d воздушного зазора полосы интерференции будут перемещаться в направлении от центра (в противоположность тому, что было с кольцами Ньютона, см. § 33, п. 7).

При увеличении зазора d на $\lambda/2$ разность хода увеличится на λ, так что произойдет смещение на одну полосу (т.е. на место каждой светлой полосы станет такая же светлая соседняя полоса). При изменении угла падения на $\Delta\varphi$ разность хода изменится на $2d\sin\varphi\,\Delta\varphi$. Отсюда видно, что полосы интерференции получатся тем шире, чем меньше d. При $d = 0$ они стали бы бесконечно широкими, т.е. поле зрения было бы освещено равномерно.

При больших зазорах и высокой степени монохроматичности света с интерферометром Майкельсона наблюдалась интерференция очень высокого порядка (около 10^6). Если M_2 и M_1' близки друг к другу и образуют воздушный клин с небольшим углом, то полосы интерференции локализуются либо на поверхности клина, либо вблизи нее. Это — полосы равной толщины, имеющие вид равноотстоящих прямых, параллельных ребру клина.

§ 36. Многолучевая интерференция

1. Вернемся к исследованию интерференции в плоскопараллельных пластинках и учтем лучи, претерпевшие в них многократные отражения (см. рис. 134). Обозначим через R коэффициент отражения света от границы раздела пластинки с воздухом, т.е. долю энергии падающего света, которая возвращается обратно при каждом отражении. При отсутствии поглощения оставшаяся доля $(1-R)$ приходит через эту границу. Если среды по обе стороны пластинки одинаковы (воздух), то, как будет показано в § 65, коэффициенты отражения на обеих поверхностях пластинки будут одинаковы. Допустим, что падающий свет монохроматический. Обозначим через I_0 его интенсивность. Тогда интенсивности прошедших пучков $1'$, $2'$, $3'$, … будут

$$I_{1'} = (1-R)^2 I_0, \quad I_{2'} = R^2(1-R)^2 I_0, \quad I_{3'} = R^4(1-R^2)I_0, \ldots,$$

а соответствующие (вещественные) амплитуды

$$a_{1'} = (1-R)a_0, \quad a_{2'} = R(1-R)a_0, \quad a_{3'} = R^2(1-R)a_0, \ldots,$$

где a_0 — амплитуда падающего света. Разность хода между двумя соседними интерферирующими пучками составляет $\Delta = 2dn\cos\psi$, а разность фаз $\Phi = k\Delta = (4\pi/\lambda)dn\cos\psi$. Амплитуда прошедшей волны представится убывающей геометрической прогрессией

$$a_d = a_0(1-R)[1 + Re^{-i\Phi} + R^2e^{-2i\Phi} + \ldots].$$

Если пластинка достаточно длинная, то прогрессию можно считать бесконечной. Тогда

$$a_d = \frac{1-R}{1-Re^{-i\Phi}}a_0.$$

Интенсивность прошедшей волны будет

$$I_d = \frac{(1-R)^2}{|1-Re^{-i\Phi}|^2}a_0^2 = \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2(\Phi/2)}I_0. \tag{36.1}$$

Рассчитаем теперь интенсивность I_r отраженной волны. Интенсивности отраженных пучков 1, 2, 3, … будут

$$I_1 = RI_0, \quad I_2 = R(1-R)^2 I_0, \quad I_3 = R^3(1-R)^2 I_0, \ldots,$$

а соответствующие (вещественные) амплитуды

$$a_1 = \sqrt{R}\,a_0, \quad a_2 = -\sqrt{R}\,(1-R)a_0, \quad a_3 = -\sqrt{R}\,R(1-R)a_0, \ldots$$

Знак минус в выражениях для амплитуд учитывает потерю полуволны при отражении на одной из поверхностей пластинки. Такую потерю полуволны не надо было учитывать при расчете амплитуды прошедшей волны, так как там все отражения происходят на границе стекло–воздух. В случае же отражения от пластинки луч 2 претерпевает отражение на границе воздух–стекло, а все остальные лучи — на границе стекло–воздух. Результирующая амплитуда отраженной волны представится геометрической прогрессией

$$a_r = \sqrt{R}\,a_0 - \sqrt{R}\,(1-R)a_0e^{-i\Phi}[1 + Re^{-i\Phi} + R^2e^{-2i\Phi} + \ldots].$$

Поступая, как раньше, найдем для интенсивности

$$I_r = \frac{4R\sin^2(\Phi/2)}{(1-R)^2 + 4R\sin^2(\Phi/2)} I_0. \tag{36.2}$$

2. Из (36.1) и (36.2) видно, что $I_d + I_r = I_0$, как это и должно быть при отсутствии поглощения. Таким образом, картины распределения интенсивностей в прошедшем и отраженном свете — *взаимно дополнительные*: светлым местам и максимумам одной картины соответствуют темные места и минимумы другой. Для получения интерференционных полос, конечно, надо пользоваться протяженными источниками света. Полосы интерференции будут *полосами равного наклона*. Их можно спроецировать линзой на экран, помещенный в ее фокальной плоскости. Можно также пользоваться зрительной трубой, установленной на бесконечность. Интерференционные полосы будут иметь форму концентрических колец, центр которых находится в точке схождения лучей, нормальных к поверхности пластинки. Максимумы в проходящем свете получаются при $\Phi/2 = m\pi$, т.е. при $2dn\cos\psi = m\lambda$, где m — целое число. В отраженном свете этим местам соответствуют минимумы.

Таким образом, положения максимумов и минимумов определяются в точности теми же условиями, что и в случае простой двухлучевой интерференции. То же относится и к распределению интенсивностей, но при том существенном условии, что коэффициент отражения мал ($R \ll 1$). Тогда в (36.1) и (36.2) можно пренебречь квадратами R и произвести разложение по R. В первом порядке получится

$$I_d = 1 - 4R\sin^2(\Phi/2) = 1 - 2R(1-\cos\Phi),$$
$$I_r = 4R\sin^2(\Phi/2) = 2R(1-\cos\Phi).$$

Но к тому же результату мы пришли бы, если бы и не учитывали многократные отражения.

Однако распределение интенсивности существенно изменяется при увеличении коэффициента отражения R, в особенности когда этот коэффициент приближается к единице. На рис. 141 приведены кривые интенсивности прошедшего света для трех значений R. Интенсивность падающего света принята за единицу. По оси абсцисс отложена разность фаз $\Phi = (4\pi dn\cos\psi)/\lambda$. При $\Phi = 2m\pi$ (m — целое) получаются максимумы. Когда $R \ll 1$, максимумы пологие. Но уже при $R = 0{,}75$ они очень резкие. При приближении R к единице весь свет практически сосредоточивается в очень узких интерференционных полосах на темном фоне. В отраженном свете получаются столь же резкие, но темные интерференционные полосы на светлом фоне. Числитель в формуле (36.1) — величина постоянная. В максимуме ($\Phi = 2m\pi$) $I_{\text{макс}} = 1$.

Резкость интерференционных полос характеризуется их *полушириной*. Для полос в проходящем свете *полушириной называется расстояние между точками, лежащими по обе стороны максимума, в которых интенсивность составляет половину максимальной величины* $I_{\text{макс}}$. В окрестности максимума

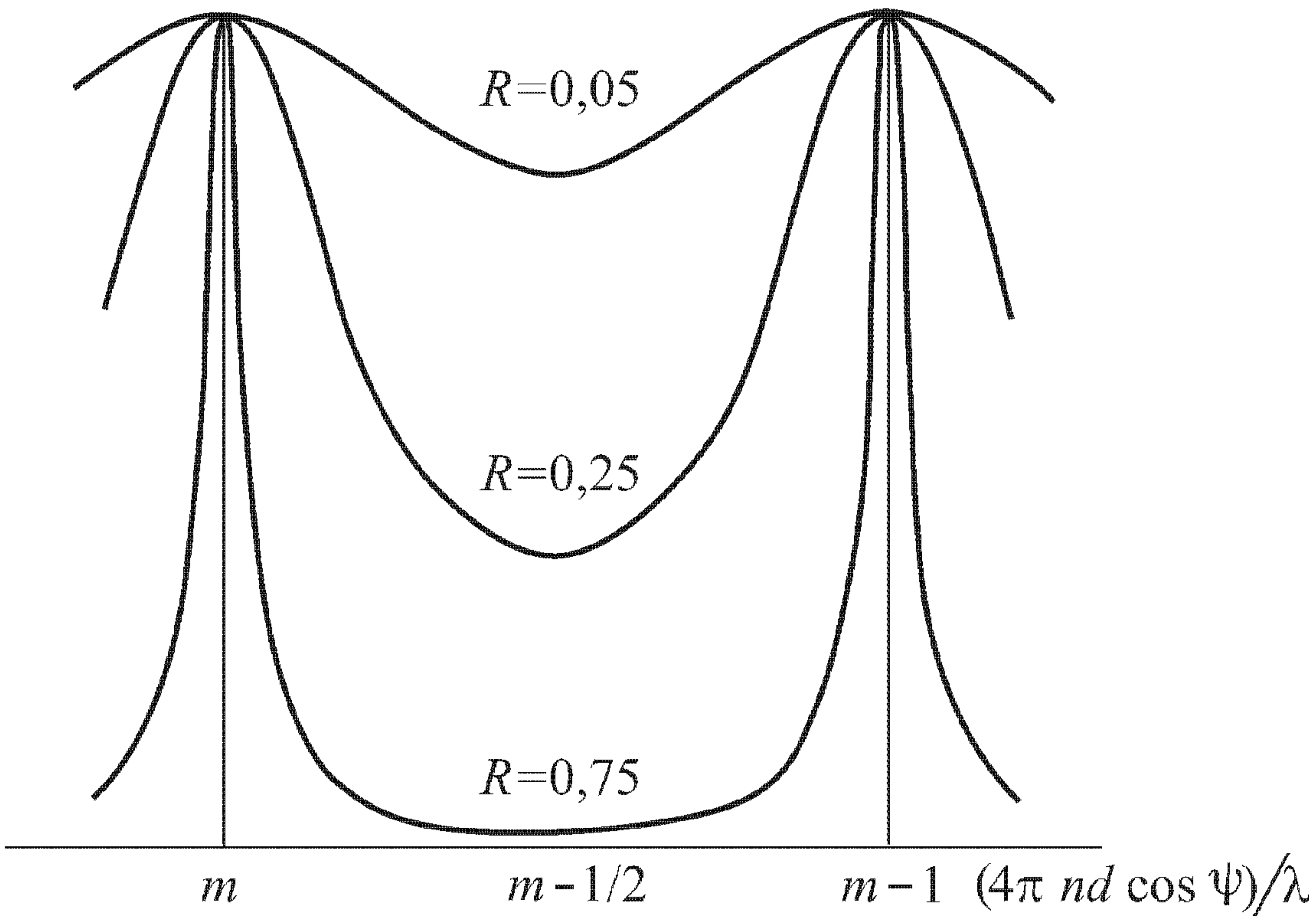


Рис. 141

m-го порядка Φ представим в виде $\Phi = m\pi + \varphi$. Тогда в формуле (36.1) Φ можно заменить на φ. Ввиду малости φ ту же формулу можно переписать так:

$$I_d = \frac{I_{\text{макс}}}{1 + R\varphi^2/(1-R)^2}. \qquad (36.3)$$

Если $R\varphi^2/(1-R)^2 = 1$, то $I_d = {}^1\!/_2 I_{\text{макс}}$. Следовательно, полуширина определится выражением

$$\delta\Phi = 2\varphi = 2\frac{1-R}{\sqrt{R}}. \qquad (36.4)$$

3. В применениях многолучевая интерференция всегда осуществляется в *толстых пластинках*, так что приходится иметь дело с интерференцией высоких порядков. Поэтому для получения интерференционных полос требуется *высокая монохроматичность света* $(\lambda/\delta\lambda \gtrsim m)$. В интерференционной спектроскопии, где многолучевая интерференция применяется для изучения спектров, исследуемый свет, если он не обладает достаточной монохроматичностью, должен быть предварительно *монохроматизован*, например с помощью призменного спектрометра или другого монохроматора.

Обычно методами интерференционной спектроскопии исследуется структура тонких спектральных линий. Достоинством интерференционных спектроскопов является их высокая разрешающая способность при большой светосиле, простоте устройства, дешевизне и удобстве в обращении. *Разрешающая способность определяет наименьшее расстояние между близкими спектральными линиями, которые изображаются в виде раздельных спектральных линий.*

Для уяснения понятия разрешающей способности допустим, что применяемый свет состоит из двух близких, одинаково интенсивных спектральных линий с длинами волн λ и $\lambda' = \lambda + \delta\lambda$. Интерференционная картина будет состоять из двух систем интерференционных полос (максимумов) с теми же длинами волн. Одна система максимумов окажется сдвинутой относительно другой. Если сдвиг недостаточен, максимумы наложатся друг на друга, получится распределение с единственным максимумом, которое для глаза неотличимо от распределения, соответствующего одной спектральной линии. Тогда говорят, что рассматриваемая двойная линия *не разрешается спектральным прибором.*

Допустим теперь, что сдвиг равен полуширине интерференционной полосы. Соответствующее распределение интенсивности показано на рис. 142. Кривые λ и λ' изображают контуры спектральных линий, как они были бы видны в спектральном приборе в отдельности. В центре картины, где контуры пересекаются, интенсивность спектральных линий составляет $^1\!/_2 I_{\text{макс}}$. При сложении интенсивностей интенсивность в центре получится равной $I_{\text{макс}}$. В точке A интенсивность первой линии равна $I_{\text{макс}}$, а второй $^1\!/_5 I_{\text{макс}}$, как это видно из формулы (36.3). Таким образом, полная интенсивность в точке A будет $I_{\text{макс}} + {}^1\!/_5 I_{\text{макс}} = 1{,}2 I_{\text{макс}}$. Такая же интенсивность получится в точке A'. Кривая результирующей интенсивности будет состоять из двух максимумов с провалом в центре между ними. Интенсивность в месте провала будет составлять $1/1{,}2 = 83\,\%$ от интенсивности одного из максимумов на результирующей кривой.

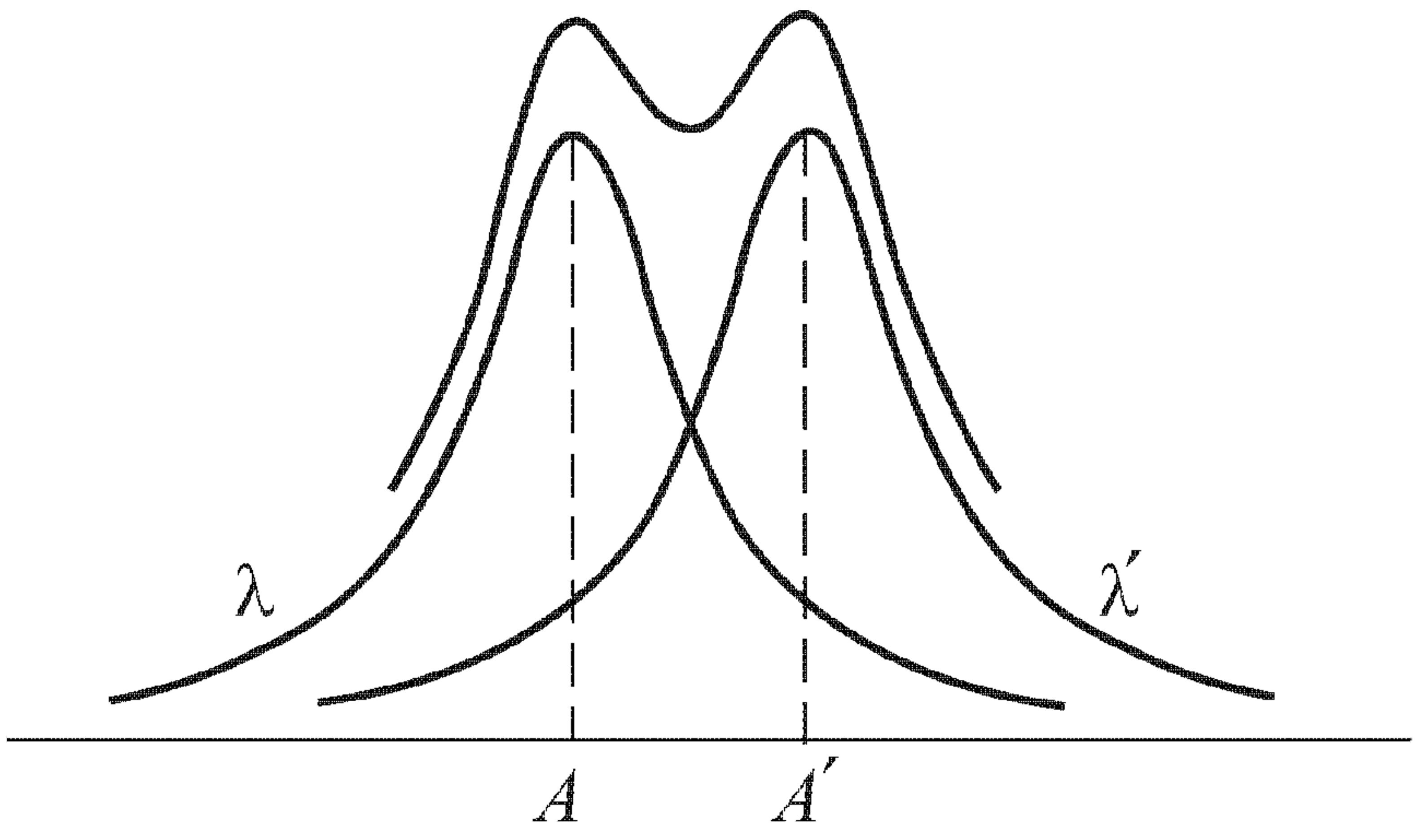


Рис. 142

Будет ли разрешаться спектральная линия при таком провале — это зависит от индивидуальных особенностей глаза. Опыт показывает, что нормальный глаз такие провалы обычно замечает без особого труда. Поэтому расстояние между максимумами спектральных линий, равное или большее полуширины интерференционной полосы, мы примем (до некоторой степени условно) за *условие спектрального разрешения*.

Определим теперь минимальную разность длин волн $\delta\lambda = \lambda' - \lambda$, соответствующую такому критерию разрешения. Все величины, относящиеся к свету с длиной волны λ, будем обозначать нештрихованными, а к длине волны λ' — штрихованными буквами. Показатель преломления n будем считать одним и тем же для обеих длин волн. (Это строго выполняется только в интерферометре Фабри–Перо.) Тогда, независимо от длины волны, всем падающим лучам одного и того же направления будут соответствовать преломленные лучи также одного направления. В точке A' для длины волны λ' получается интерференционный максимум m-го порядка, а потому $\Phi' = 2m\pi$. В той же точке волна с длиной λ максимума уже не дает. Для такой волны разность фаз имеет значение $\Phi = 2m\pi + (1-R)/\sqrt{R}$, т.е. в рассматриваемой точке $\Phi' - \Phi \equiv \delta\Phi = (1-R)/\sqrt{R}$. Но ввиду одинаковости n и ψ для обеих волн из формулы $\Phi = (4\pi dn\cos\psi)/\lambda$ следует: $\delta\Phi/\Phi = |\delta\lambda/\lambda|$. Учтя, что в максимуме $\Phi = 2\pi m$, поэтому находим

$$\frac{\lambda}{\delta\lambda} = \frac{2\pi\sqrt{R}}{1-R}m. \tag{36.5}$$

Величина $\lambda/\delta\lambda$ называется *разрешающей способностью спектрального прибора*.

Интерференционные спектральные аппараты осуществлены в виде двух приборов, которые мы и рассмотрим.

4. Интерферометр или эталон Фабри–Перо. Этот прибор, изобретенный Фабри (1867–1945) и Перо (1863–1925), является наиболее распространенным из интерференционных спектральных приборов. Он состоит из двух стеклянных или кварцевых пластинок P_1 и P_2, между которыми обычно находится воздух (рис. 143). Плоские поверхности пластинок, обращенные друг к другу, тщательно отшлифованы и покрыты высокоотражающими слоями (серебро, алюминий, многослой-

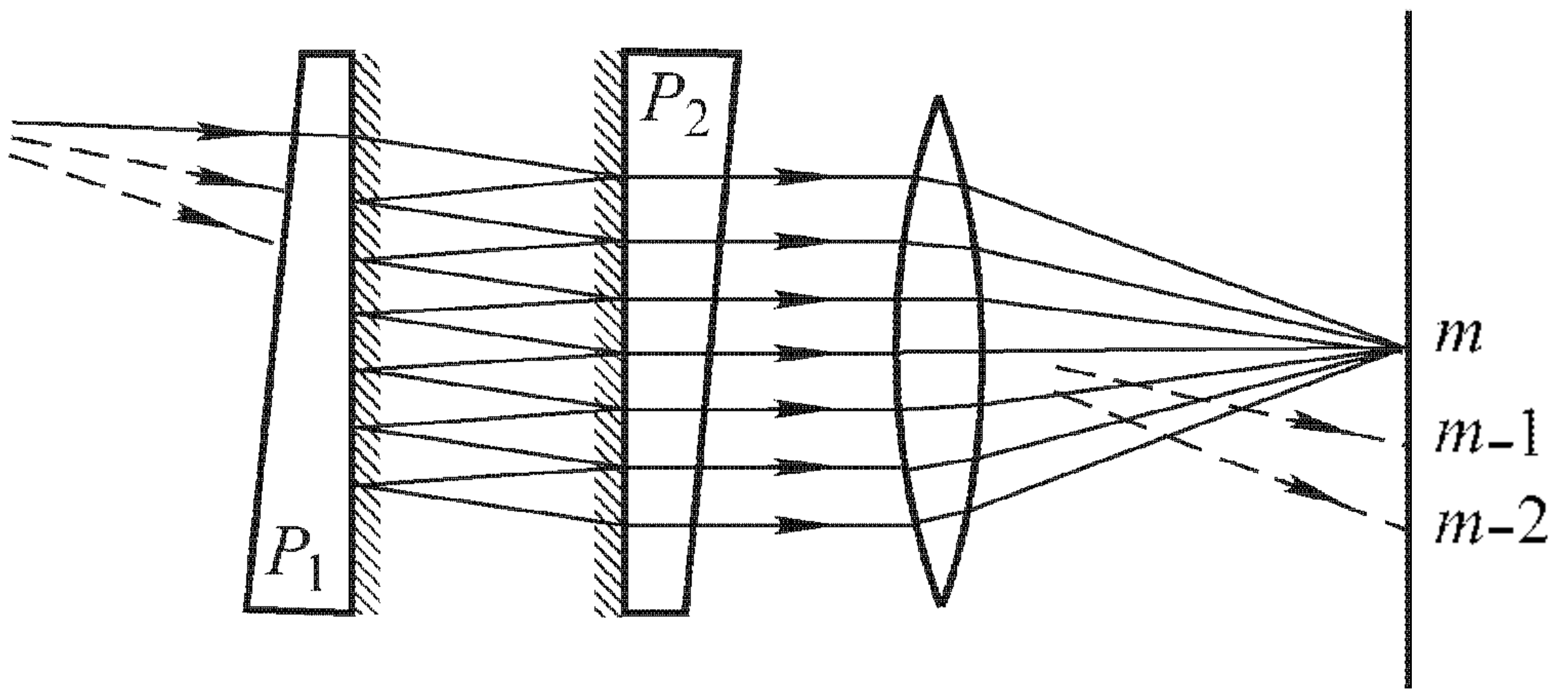


Рис. 143

ные диэлектрические покрытия). В хороших приборах отступления внутренних поверхностей пластинок от идеальных плоскостей не превышают $0{,}01\lambda$. Отражательная способность зеркальных поверхностей пластинок при металлическом покрытии может быть доведена до 95 %, а при диэлектрических — до 98 %.

Параллельность зеркальных поверхностей достигается с помощью распорного кольца из инвара или плавленого кварца, помещаемого между пластинками. Кольцо снабжено тремя выступами с каждой стороны, к которым пластинки прижимаются при помощи трех пружин. Выступы подшлифованы так, что зеркальные поверхности устанавливаются параллельно друг другу. Небольшие отступления от параллельности устраняются нажимом соответствующей пружины.

Интерферометр, в котором используются распорные кольца, называется *эталоном Фабри–Перо.* Располагая набором эталонов с кольцами разной толщины, можно производить измерения при различных расстояниях между зеркальными поверхностями. В прежних моделях интерферометра Фабри–Перо можно было с помощью специальных микрометрических винтов менять расстояние между зеркалами. Но таким путем не удается достигнуть той высокой точности, которую можно получить с помощью эталона. Наружные поверхности пластинок обычно образуют небольшие углы с внутренними, чтобы отраженный от них светлый блик не мешал наблюдению основной интерференционной картины. Таким образом, интерферометр Фабри–Перо можно рассматривать как плоскопараллельную воздушную пластинку, на которой происходят многократные отражения световых лучей и последующая интерференция их.

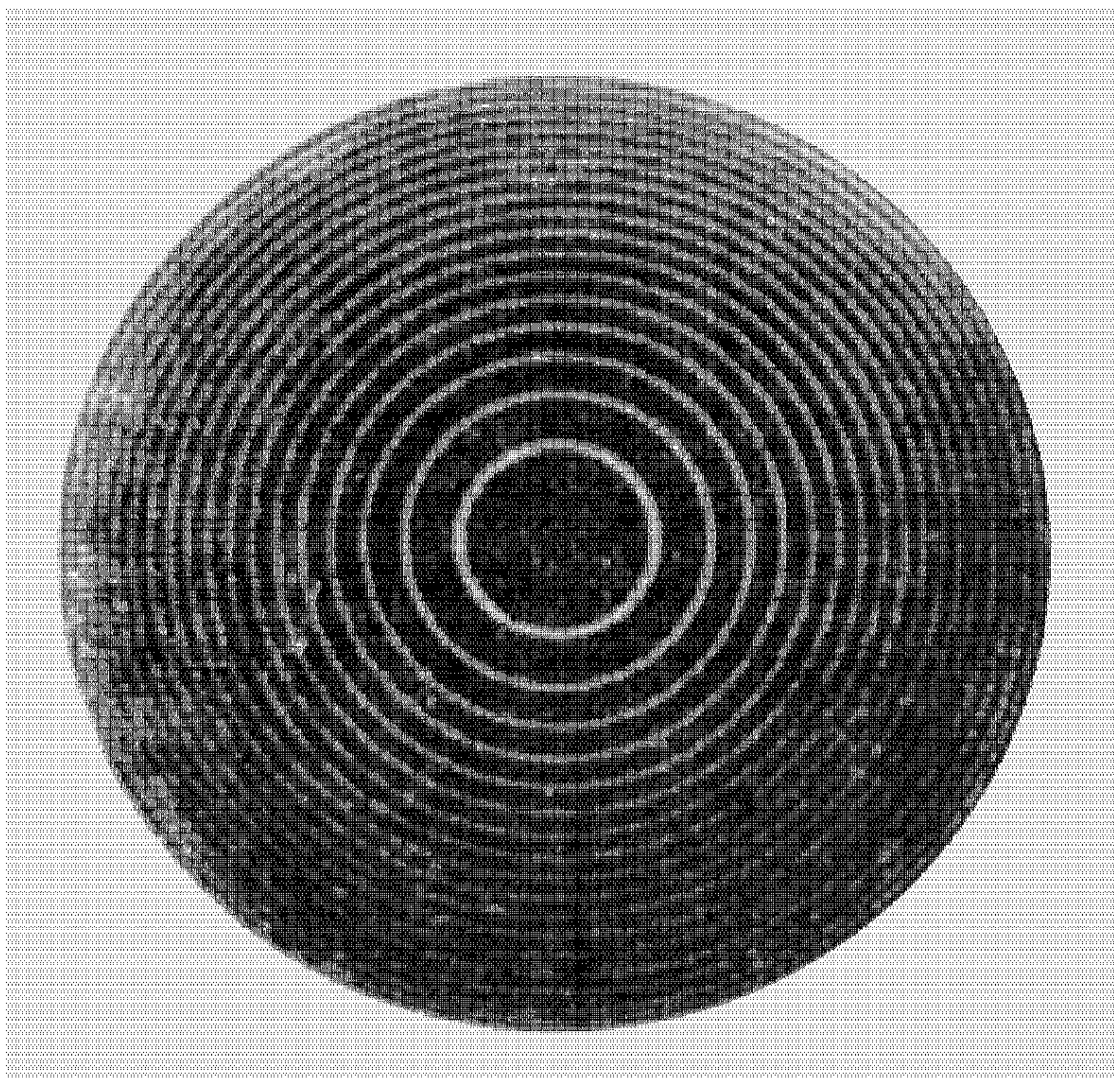


Рис. 144

Интерференционная картина состоит из концентрических колец равного наклона. Она приведена на рис. 144 в проходящем свете. Интерференционные максимумы тем у́же, чем больше отражательная способность зеркальных поверхностей пластинок интерферометра.

Расстояние между зеркалами h обычно составляет 1–100 мм, а в специальных эталонах доходит до 1 м. Поэтому порядки интерференции $m \approx 2h/\lambda$ очень высоки (при $h = 5$ мм $m \approx 20\,000$).

Ввиду малости угла ψ условие главного интерференционного максимума $2h\cos\psi = m\lambda$ можно записать в виде $h(2-\psi^2) = m\lambda$. Отсюда находим *угловую дисперсию* интерферометра Фабри–Перо:

$$\frac{d\psi}{d\lambda} = -\frac{m}{2h\psi} = -\frac{1}{\lambda\psi}. \tag{36.6}$$

При рабочих условиях ($\psi \approx 10^{-2}$ рад) угловая дисперсия интерферометра Фабри–Перо значительно превышает дисперсию других спектральных аппаратов. Это является основным преимуществом интерферометра Фабри–Перо.

Другим важным преимуществом интерферометра Фабри–Перо является его большая светосила. Благодаря этому и своей дешевизне такой интерферометр получил широкое распространение для спектральных исследований в оптической области спектра. Принцип интерферометра Фабри–Перо был использован также в объемных резонаторах оптических квантовых генераторов (см. § 120).

5. Пластинка Луммера–Герке (1878–1960). Она представляет собой плоскопараллельную пластинку из очень однородного стекла или плавленого кварца толщиной от 3 до 10 мм и длиной до 30 см. Для направления световых лучей в пластинку на одном конце ее сбоку посажена на оптический контакт добавочная призмочка (рис. 145 *а*). При другом способе один конец пластинки скошен (рис. 145 *б*). В обоих случаях падающие лучи нормальны к поверхности стекла, чем достига-

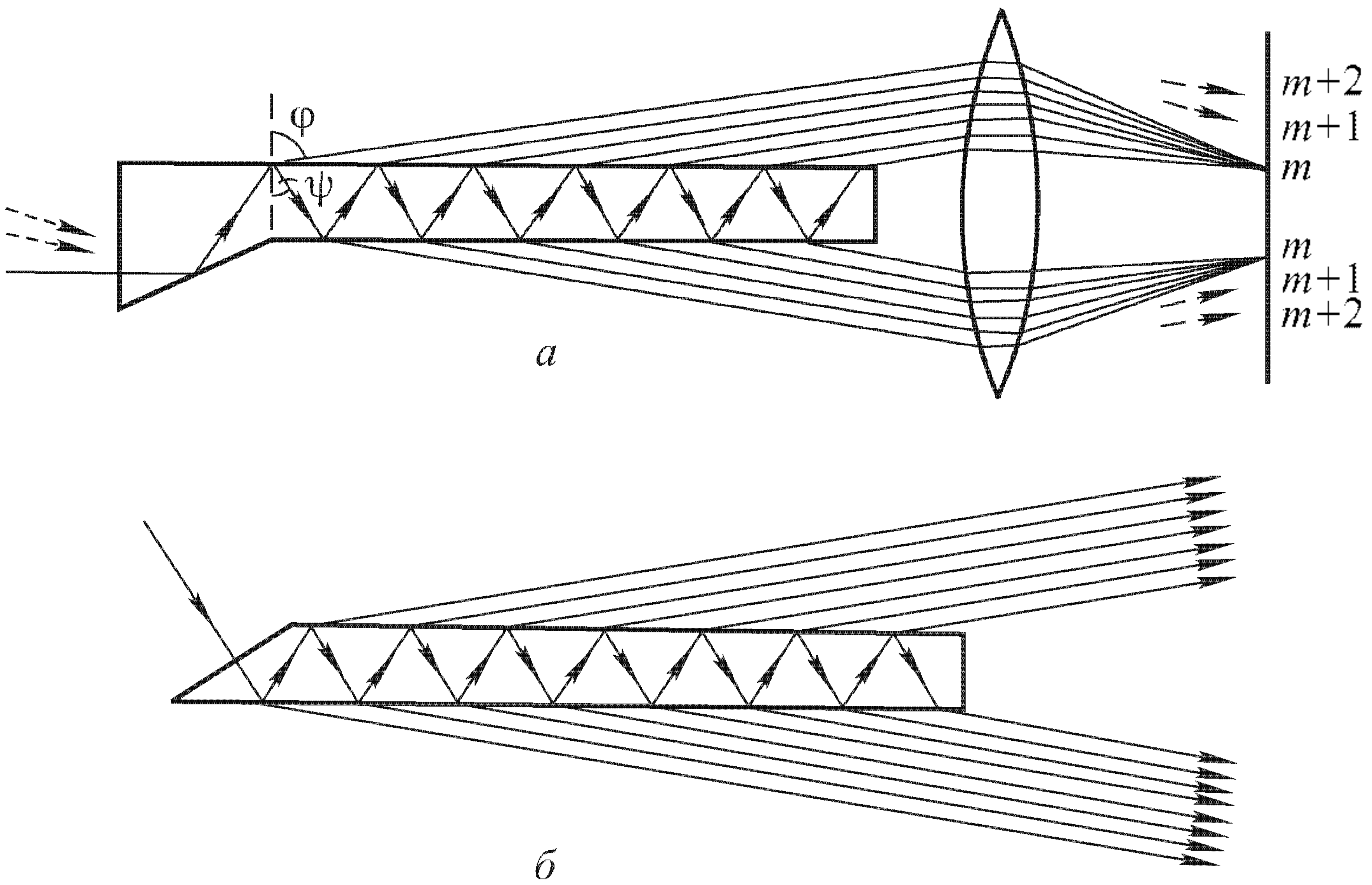


Рис. 145

ется уменьшение потерь света на отражение. Направление падающих лучей подбирается таким, чтобы угол падения на границе стекло–воздух был близок к предельному углу полного отражения. Тогда коэффициент отражения мало отличается от единицы. Пучки испытывают многократные отражения от плоскостей пластинки и выходят из нее с почти одинаковыми интенсивностями. Можно получить до 10–15 таких пучков с каждой стороны пластинки.

В пластинке Луммера–Герке наблюдаются интерференционные полосы равного наклона. Условие интерференционного максимума m-го порядка имеет вид

$$2hn\cos\psi = m\lambda, \tag{36.7}$$

где h — толщина пластинки, а ψ — угол преломления. Здесь не учтено изменение фазы при отражении, так как оно производит только несущественное смещение всей интерференционной картины. Число интерферирующих пучков по одну сторону пластинки равно $N = L/(2h\,\mathrm{tg}\,\psi)$, где L — длина пластинки. А так как свет падает почти под предельным углом, то

$$N = \frac{L}{2h}\sqrt{n^2-1}. \tag{36.8}$$

Вопрос о разрешающей способности пластинки Луммера–Герке будет разобран в § 48.

6. Затронем еще два вопроса, имеющих отношение к многолучевой интерференции. Рассмотрим сначала много равноотстоящих параллельных отражающих плоскостей с малыми коэффициентами отражения (рис. 146). Почти весь свет, падающий на каждую плоскость, проходит через нее без потерь. В результате получится много параллельных отраженных лучей *1*, *2*, *3*, … почти одинаковой интенсивности. Разность хода между соседними лучами составляет $2d\cos\varphi$, где d — расстояние между соседними плоскостями. Если выполнено условие

$$2d\cos\varphi = m\lambda \tag{36.9}$$

(m — целое число), то все лучи будут находиться в фазе. Возникнет сильное отражение, называемое *интерференционным отражением*. При заданном m условие (36.9) может выполняться только для определенных длин волн. Поэтому при освещении рассматриваемой системы плоскостей белым светом интерференционное отражение будет наблюдаться лишь для некоторых (а при $d < \lambda$ — для одной) сравнительно узких спектральных линий. Система плоскостей, таким образом, при-

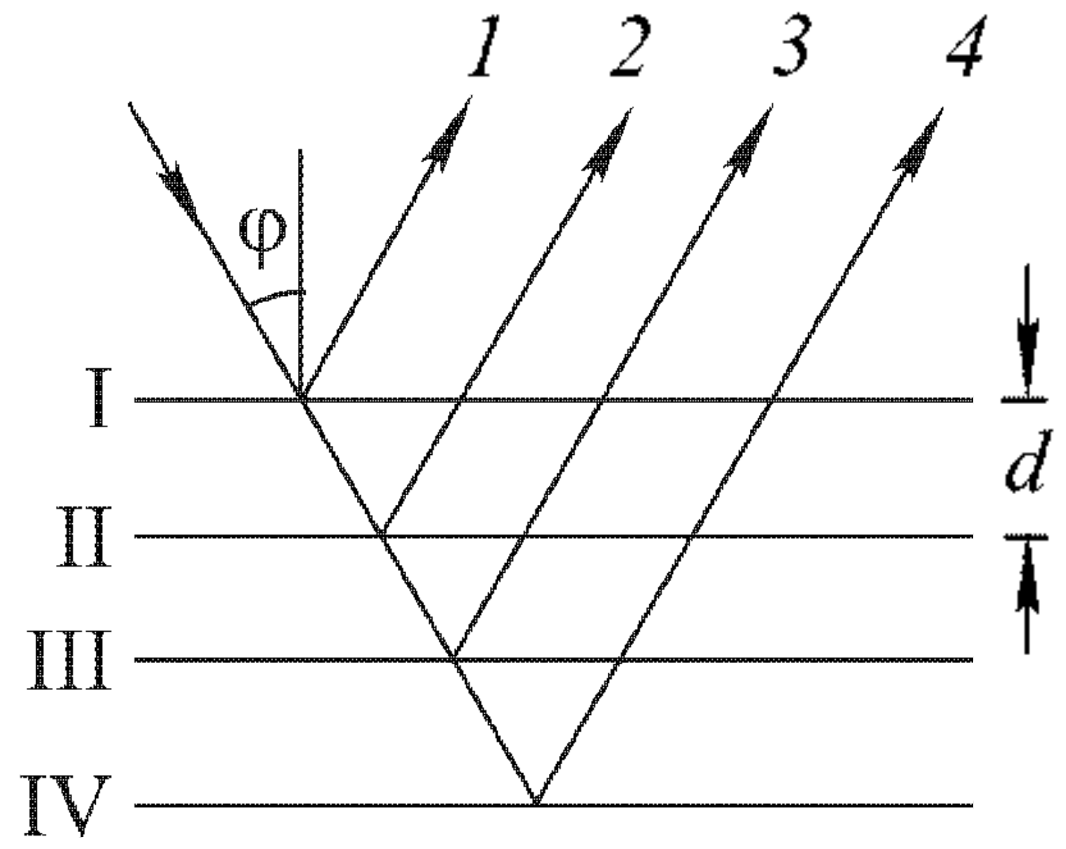


Рис. 146

обретает свойство *избирательного отражения.* При этом, так как разность хода зависит от угла падения, цвет избирательно отраженного света будет меняться с изменением угла падения.

Отметим также, не входя ни в какие подробности, что надлежащей комбинацией диэлектрических слоев с различными показателями преломления удается осуществить на поверхности стекла *многослойные покрытия* (пленки) с высокими коэффициентами отражения (до 98–99 % для определенных длин волн). Такие покрытия применяются, например, в зеркалах интерферометров Фабри–Перо. Следует, однако, иметь в виду, что отражающие многослойные пленки в белом свете окрашены, так как интерференционное условие сильного отражения точно выполняется только для какой-либо одной длины волны.

§ 37. Стоячие световые волны

1. При интерференции монохроматических волн, распространяющихся навстречу друг другу, образуются *стоячие волны* (см. т. III, § 140). В этом случае, как видно из формулы (26.12), ширина полосы Δx равна $\lambda/2$, т.е. расстоянию между соседними пучностями или узлами. В случае электромагнитных волн пучности (узлы) электрического вектора $\mathbf{E}$ совпадают с узлами (пучностями) магнитного вектора $\mathbf{B}$. Таким образом, в стоячей волне электрическое и магнитное поля можно *пространственно разделить* и исследовать свойства и действия этих полей в отдельности. Моменты прохождения электрического и магнитного полей через максимум в стоячей волне не совпадают, а сдвинуты относительно друг друга на половину периода световых колебаний. Получение стоячих световых волн наталкивается на трудности, связанные с малостью длины волны. Эти трудности впервые были преодолены О. Винером (1862–1927) в 1890 г.

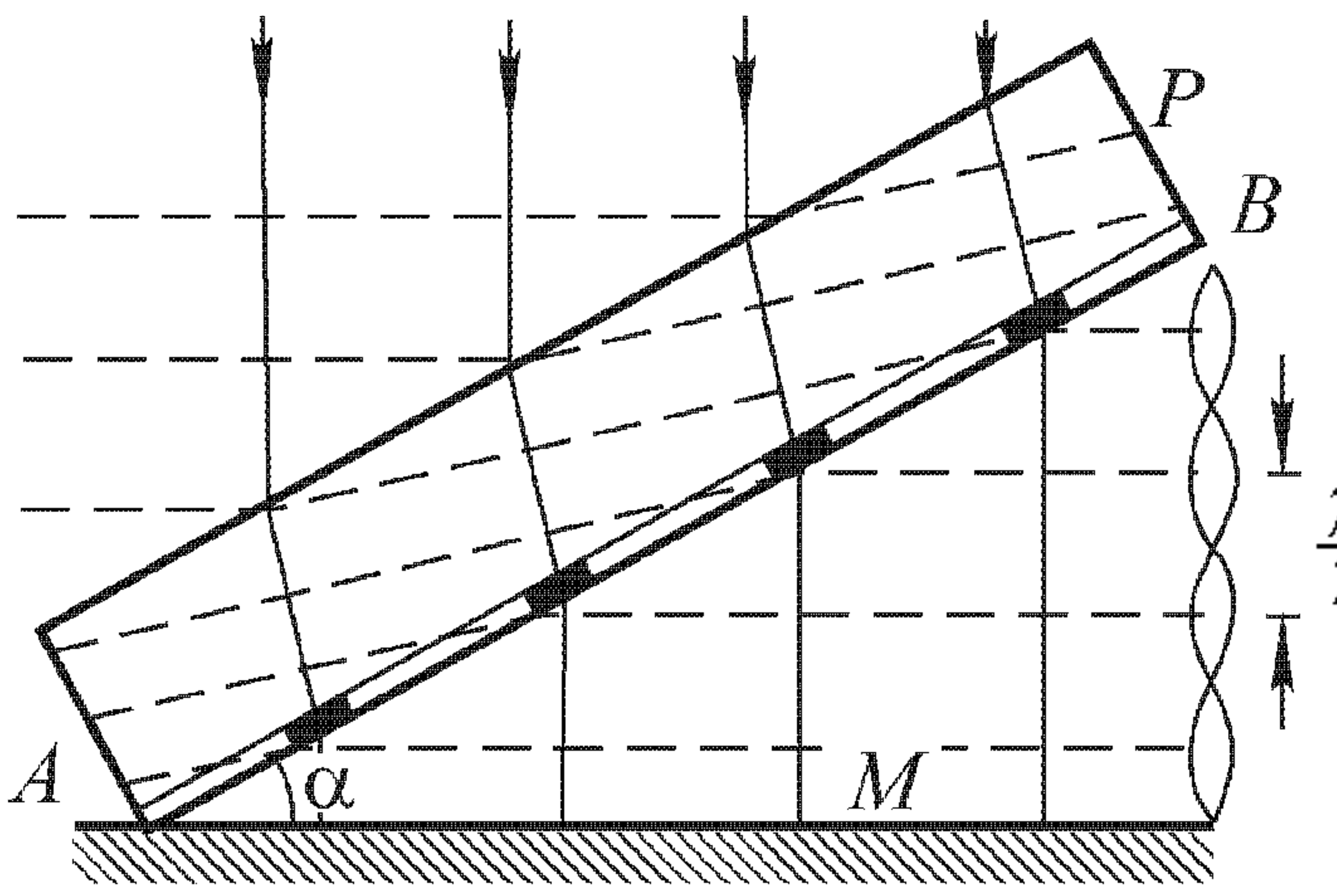


Рис. 147

Стеклянная пластинка P (рис. 147), покрытая тонким фоточувствительным слоем толщиной $\sim \dfrac{\lambda}{20} - \dfrac{\lambda}{30}$, приводилась в соприкосновение с хорошим металлическим зеркалом M, образуя с ним очень малый угол α (порядка $1'$ или меньше). Пластинка освещалась монохроматическими лучами, нормальными к поверхности зеркала.

В результате отражения от зеркала в воздушном клине между зеркалом и фотослоем образовывались стоячие световые волны. Их пучности обнаруживались в фотослое после его проявления в виде равноотстоящих почерневших полос, параллельных ребру воздушного клина. Расстояние между этими полосами равно $\lambda/(2\sin\alpha)$. Полосы можно видеть глазом, если угол α взять достаточно малым. В опытах Винера расстояние между полосами составляло 1–2 мм. По этим данным можно было оценить длину световой волны. Удадось также показать, что для красного света длина волны примерно вдвое больше, чем для синего.

Выдвигалось возражение, что при интерпретации опыта не было учтено обратное отражение света от фотослоя. Чтобы снять это возражение, Винер заполнял пространство между фотослоем и металлическим зеркалом бензолом, показатель преломления которого близок к показателю преломления желатины фотослоя. Тем самым устранялось влияние отражения от фотослоя. Оказалось, что при этом вид интерференционных полос не менялся.

Главное значение опытов Винера состоит в том, что они позволили установить, какой вектор производит фотохимическое действие: электрический или магнитный. Опыты показали, что первый слой почернения получается на расстоянии $\lambda/4$ от поверхности металлического зеркала. Здесь поэтому находится пучность вектора, вызывающего фотографическое действие. Но при отражении электромагнитной волны от металла на его поверхности получается узел электрического вектора (см. т. III, § 45). Первая пучность электрического вектора находится на расстоянии $\lambda/4$ от поверхности металлического зеркала — там же, где наблюдается почернение фотографического слоя. Отсюда следует, что фотографическое (и вообще фотохимическое) действие производится электрическим, а не магнитным полем световой волны.

Позднее, в 1892 г., Друде (1863–1906) и Нернст (1864–1941) повторили опыт Винера, заменив фотографический слой тонкой пленкой флуоресцирующего вещества. Айвс в 1933 г. сделал то же самое с фотоэлектрическим слоем. Оказалось, что флуоресценция и фотоэффект вызываются также электрическим полем. Этого и следовало ожидать с точки зрения электронных представлений о строении вещества.

Действительно, действие света на вещество сводится к силам, с которыми на электроны действуют электрические и магнитные поля. Первая сила равна $e\mathbf{E}$, вторая $\frac{e}{c}[\mathbf{vB}]$, где $\mathbf{v}$ — скорость электрона. Магнитная сила много меньше электрической, так как $v \ll c$. Таким образом, во всех действиях электрический вектор световой волны является более важным, чем магнитный. Поэтому его нередко называют *световым вектором.*

2. Винер исследовал также стоячие световые волны в поляризованном свете. При отражении от прозрачной среды под определенным углом отраженный свет оказывается линейно поляризованным (см. § 65). Про такой свет теперь говорят, что он *поляризован перпендикулярно к плоскости падения.* Со времен Френеля (1821 г.) до конца XIX века во всех эфирных теориях света обсуждался вопрос, как направлен световой вектор: параллельно плоскости поляризации или перпендикулярно к ней. Опыт Винера и должен был решить этот вопрос.

Винер изменил постановку своего предыдущего опыта, заставив линейно поляризованный свет отражаться от металлического зеркала под углом 45°. Тогда угол между направлениями распространения падающей и отраженной волн будет 90°. Поэтому при интерференции этих волн расстояние между соседними пучностями или узлами стоячей волны окажется равным $\lambda/\sqrt{2}$, как это видно из формулы (26.12). Если световой вектор перпендикулярен к плоскости падения, то колебания в падающей и отраженной волнах будут параллельными. Тогда возникнет интерференция этих волн и на фотографической пластинке, наклоненной к поверхности металлического зеркала, получатся такие же темные и светлые полосы, как и в первом опыте Винера. Если же световой вектор параллелен плоскости падения, то интерференция невозможна, так как в этом случае колебания в падающем и отраженном свете будут взаимно перпендикулярны (см. § 26, п. 5).

Опыт показал, что полосы получаются, когда свет поляризован перпендикулярно к плоскости падения. Если же он поляризован в плоскости падения, то получается равномерное почернение пленки. Отсюда следует, что *световой вектор лежит в плоскости поляризации.* А так как предыдущими опытами доказано, что он является электрическим вектором, то последний также лежит в плоскости поляризации.

3. На явлении стоячих волн основан метод *цветной фотографии*, разработанный Липпманом (1845–1922) в 1891 г. Стеклянная пластинка покрыта с одной стороны толстым слоем мелкозернистой сухой фотографической эмульсии высокой чувствительности. Пластинка составляет стенку сосуда, в который наливается ртуть. Эмульсией пластинка должна быть обращена внутрь сосуда, чтобы фотографический слой находился в контакте с ртутью. Пластинка освещается снаружи нормально падающими лучами определенной длины волны. При отражении света от поверхности ртути в толще эмульсии образуются стоячие волны, которые и экспонируют фотопластинку. После этого пластинка вынимается, проявляется и фиксируется. В местах пучностей электрического вектора выделяются тонкие слои металлического серебра, находящиеся на расстоянии $\lambda/2$ друг от друга. Получается плоскослоистая среда из отражающих параллельных плоскостей, подобная той, которая изображена на рис. 146.

Осветим теперь со стороны эмульсии так обработанную пластинку нормально падающим светом. Малая часть света отразится от слоя I, остальная часть пройдет прямо. От второго слоя отразится столь же малая часть, и т.д. Получится столько же отраженных пучков, сколько металлических слоев отложилось в толще эмульсии. Интенсивность каждого пучка мала, но их число велико. Если длины волн освещающего света и света, применявшегося при экспонировании, в точности одинаковы, то каждый последующий отраженный пучок будет запаздывать по фазе относительно предыдущего на 2π. В этом случае произойдет интерференционное усиление пучков, так что результирующая интенсивность отраженного света окажется большой. Если же каждый пучок будет сдвинут по фазе относительно предыдущего на $\pm 2\pi/N$, где N — общее число пучков, то произойдет их взаимное гашение. Вообще, интенсивно будут отражаться только такие длины волн, которые равны длине волны света, применявшегося при экспонировании, или очень близки к ней. Поэтому при освещении пластинки белым светом она в отраженном свете будет казаться окрашенной в соответствующий цвет.

Спектральная область $\delta\lambda$ отраженного света тем у́же, чем толще слой эмульсии, чем больше число отложившихся слоев металлического серебра. Если на эмульсию подышать, то она немного разбухнет, расстояния между металлическими слоями увеличатся, что вызовет смещение цвета в красную сторону. При увеличении угла падения φ условие интерференционного усиления $2d \sin\varphi = \lambda$ будет выполняться для более коротких волн. Поэтому при наклоне пластинки ее цвет меняется в сторону фиолетового конца спектра.

Для получения цветных фотографий предметов при экспонировании пластинки надо получать цветное изображение фотографируемого предмета на светочувствительном слое, как во всяком фотоаппарате. Липпмановские цветные фотографии получаются высокого качества в чистых спектральных цветах, например при фотографировании спектров. Смешанные цвета, как правило, дают плохие результаты. Метод Липпмана не получил распространения. Современная техническая цветная фотография основана на принципе светофильтров — использовании соответствующих красящих пигментов, вводимых в эмульсию фотопластинок.

Нейхаусс в конце 90-х годов прошлого века на опыте убедился в слоистой структуре пленок Липпмана. Была получена по методу Липпмана фотография в красном свете. После этого пленка была отделена от стекла и произведен поперечный разрез ее. Затем была получена микрофотография одного из разрезов пленки с увеличением в 1000 раз. На микрофотографии было видно около десяти темных полос, чередовавшихся со светлыми. Темные полосы получились в местах отложения металлического серебра. Расстояние между серединами темных полос получилось ≈ 350 нм. Необходимо заметить, что никакой микроскоп не позволяет различить детали предмета, много меньшие длины световой волны. Успех опыта отчасти определялся тем, что пленка заметно разбухла при обработке до ее разрезания. Кроме того, разрез производился в косом направлении.

§ 38. Излучение Вавилова–Черенкова

1. Интерференция света — отнюдь не редкое явление, как это может показаться с первого взгляда. По существу все, что связано с распространением света в веществе, относится к классу явлений интерференции. Действительно, световая волна, падая на среду, возбуждает электрические колебания в атомах и молекулах. Вследствие этого эти частицы начинают излучать вторичные электромагнитные волны, также воздействующие на атомы и молекулы среды. Падающая волна и возбужденные ею вторичные волны интерферируют, и этой интерференцией определяются все особенности распространения света в среде.

Действие линзы и вогнутого зеркала можно рассматривать также как интерференционный эффект. Действительно, возмущения от точечного источника, если они попадают на различные участки линзы, в дальнейшем распространяются по различным путям вдоль лучей к фокусу. Так как оптические длины всех лучей от источника до фокуса одинаковы, то возмущения приходят в фокус в одинаковых фазах и при наложении усиливают друг друга. Этим интерференционным эффектом и объясняется фокусирующее действие линзы. Если бы при своем распространении световые возмущения строго следовали законам лучевой (геометрической) оптики, то в фокусе получилась бы бесконечная интенсивность светового поля. То обстоятельство, что этого не происходит, связано с отступлениями от геометрической оптики и прежде всего с дифракцией.

2. Рассмотрим более интересный пример. Допустим, что в прозрачной однородной среде движется электрон или какая-либо другая заряженная частица с постоянной скоростью V. Своим полем движущийся электрон возбуждает атомы и молекулы среды, и они становятся центрами излучения электромагнитных волн. При равномерном движении электрона эти волны *когерентны* и могут интерферировать между собой. Если скорость электрона V больше фазовой скорости света в среде v, то волны, исходящие от электрона в различные моменты времени, при определенных условиях могут приходить в точку наблюдения одновременно.

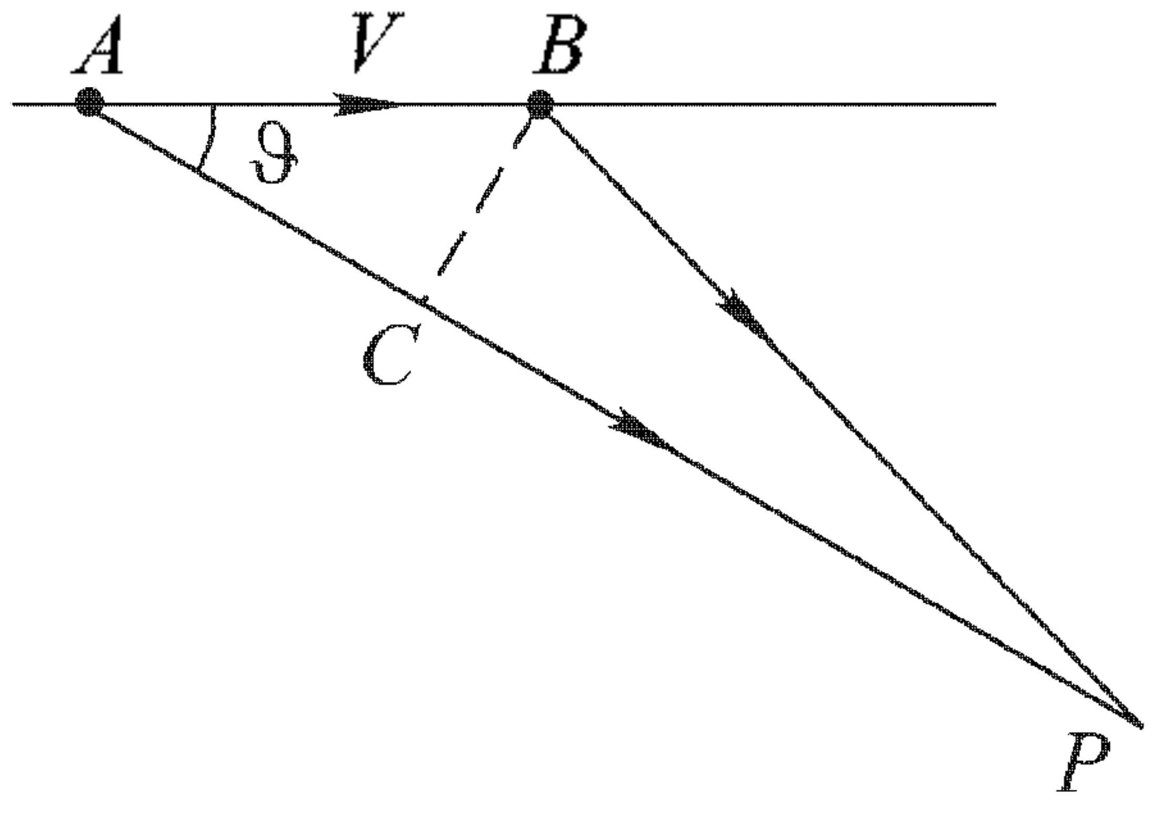


Рис. 148

Действительно, пусть A и B (рис. 148) — точки, через которые электрон проходил в моменты времени t_1 и t_2 соответственно. На прохождение расстояния AB электрон затратил время $t_2 - t_1 = AB/V$. В точку наблюдения P волны из A и B придут в моменты времени $t_1 + AP/v$ и $t_2 + BP/v$. Разность этих времен равна

$$\Delta t = (t_2 - t_1) - \frac{AP - BP}{v} = \frac{AB}{V} - \frac{AP - BP}{v}.$$

Если точка P удалена достаточно далеко, то $AP - BP \approx AC = AB\cos\vartheta$, так что $\Delta t = AB\left(\frac{1}{V} - \frac{\cos\vartheta}{v}\right)$. Так как по предположению $v < V$, то существует угол ϑ, удовлетворяющий условию

$$\cos\vartheta = \frac{v}{V} = \frac{1}{n\beta}, \tag{38.1}$$

где $\beta \equiv V/c$, а $n \equiv c/v$ — показатель преломления среды. Если это условие выполнено, то все волны придут в точку P *одновременно*, какова бы ни была длина отрезка AB. В этом случае при интерференции произойдет взаимное усиление их. Во всех других случаях путь электрона можно разбить на такие отрезки, чтобы от крайних точек каждого из них волны приходили в точку P с разностью хода λ. Волны, приходящие от всех точек каждого из таких отрезков, полностью погасят друг друга из-за интерференции. Следовательно, то же произойдет с волнами, приходящими от всех точек среды, лежащих на пути движения электрона. Таким образом, в направлениях, определяемых условием (38.1), электрон (точнее — среда, в которой он движется) будет излучать электромагнитные волны, а в остальных направлениях излучения не будет.

Такое излучение экспериментально было обнаружено в 1934 г., а затем подробно исследовано Ф.А. Черенковым (р. 1904), в то время аспирантом С.И. Вавилова (1891–1951). Черенков показал, что все без исключения жидкие и твердые тела при прохождении через них быстрых электронов, помимо флуоресценции, имеющей место в некоторых случаях, всегда испускают слабый видимый свет с непрерывным спектром. Свечение частично поляризовано, причем электрический вектор лежит преимущественно в плоскости, образованной световым лучом излучения и направлением движения электрона. Излучение направлено главным образом вперед и максимально вдоль образующих конуса, ось которого задается направлением движения электрона, а угол при вершине 2ϑ определяется формулой (38.1). Свечение не удается потушить ни температурным воздействием, ни прибавлением к светящейся среде веществ, тушащих флуоресценцию. Попытки определения времени нахождения излучающих атомов и молекул в возбужденном состоянии показали, что это время равно нулю: свечение прекращается одновременно с прекращением прохождения электронов через среду. Эти особенности свечения привели С.И. Вавилова к заключению, что оно не является люминесценцией, для которой характерна конечная длительность свечения.

Ввиду очень слабой яркости свечения, Черенков при количественных измерениях в большинстве случаев применял метод Вавилова фотометрирования по порогу зрения.

В дальнейшем было показано, что свечение вызывается также протонами, мезонами и другими быстрыми заряженными частицами. Свечение, вызываемое радиоактивными излучениями, было известно и ранее, но оно во всех случаях неправильно считалось одним из видов люминесценции.

3. Ведущая роль в руководстве экспериментальными исследованиями, направленными на выяснение природы свечения, принадлежит С.И. Вавилову. Качественное объяснение явления излучения Вавилова–Черенкова, приведенное выше, было дано И.Е. Таммом (1895–1971) и И.М. Франком (р. 1908) в 1934 г. Тогда же ими была создана и количественная теория, согласующаяся с наблюдаемыми фактами. Несколько позднее, в 1940 г., В.Л. Гинзбург (р. 1916) построил квантовую теорию, основанную на законах сохранения энергии и импульса.

Из рассуждений, приведенных выше, ясно, что в формулу (38.1) должна входить *фазовая*, а не какая-либо другая скорость света, так как именно она определяет фазы колебаний, а с ними и условие интерференционного усиления волн. Из формулы (38.1) видно, что излучение на частотах ω, для которых $n(\omega) < 1/\beta$, невозможно. Поэтому спектр излучения Вавилова–Черенкова должен обрываться на коротких волнах, где условие (38.1) перестает выполняться из-за дисперсии света. В частности, невозможно излучение Вавилова–Черенкова рентгеновских волн, так как для них $n < 1$.

Если бы электрон двигался в среде строго равномерно, то излучение было бы точно сосредоточено на поверхности конуса, определяемого условием (38.1). Из-за неравномерности движения появляется размытие этой поверхности. Но даже и при наличии такого размытия направленность излучения указывает на то, что молекулы и атомы среды, возбуждаемые электроном, излучают *когерентно* если и не на всем пути движения электрона, то во всяком случае на его части порядка длины световой волны. Следовательно, рассматриваемое явление *практически не зависит от атомного строения среды* и может быть рассмотрено в рамках *макроскопической электромагнитной теории.* Это и было сделано Таммом и Франком. Ограничимся приведением основного результата. Полная энергия, излучаемая электроном в единицу времени на единице его пути, определяется формулой

$$\frac{dW}{dt} = \frac{e^2}{c^2} \int\limits_{\beta n > 1} \left(1 - \frac{1}{\beta^2 n^2}\right) \omega \, d\omega. \tag{38.2}$$

Интегрирование производится по частотам, для которых $\beta n(\omega) > 1$. Той же формулой очевидным образом определяется и распределение излучаемой энергии по спектру частот. Формула справедлива при выполнении условия

$$T(\omega) \left|\frac{dV}{dt}\right| \ll v(\omega), \tag{38.3}$$

т.е. когда скорость электрона V за период световых колебаний $T = 2\pi/\omega$ меняется мало по сравнению с фазовой скоростью $v(\omega)$. По оценке Тамма и Франка полная потеря энергии электрона на излучение Вавилова–Черенкова в жидкостях и твердых телах по порядку величины составляет несколько тысяч электрон-вольт на сантиметр пути, т.е. пренебрежимо мала по сравнению с потерями энергии по другим причинам.

Излучение электрона, конечно, приводит к его торможению. Само по себе ускоренное движение электрона вызывает излучение. Однако из изложенного выше следует, что это излучение не имеет ничего общего с интерференцией, определяющей излучение Вавилова–Черенкова. Если бы к электрону приложить силу, уравновешивающую все тормозящие силы, то ускорение исчезло бы, а излучение Вавилова–Черенкова осталось бы. Именно так надо понимать утверждение, что электрон, равномерно движущийся в среде, излучает, если его скорость больше фазовой скорости света в этой среде.

4. Явления, аналогичные излучению Вавилова–Черенкова, давно известны в гидро- и аэродинамике. Например, если скорость судна превышает минимальную фазовую скорость на поверхности воды, то оно начинает непрерывно генерировать волновые цуги, даже когда его скорость остается постоянной. В этом причина так называемого *волнового сопротивления*, встречаемого судами при движении по поверхности воды. То же самое происходит при полете снарядов и самолетов в воздухе. Они начинают излучать ударные волны, называемые *волнами Маха* (1838–1916), если движутся со сверхзвуковыми скоростями. Потери энергии на это излучение столь значительны, что они обусловливают основное сопротивление, испытываемое снарядами и самолетами при сверхзвуковых движениях. Указанные явления сложнее излучения Вавилова–Черенкова из-за нелинейности гидродинамических уравнений.

5. Излучение Вавилова–Черенкова может вызываться не только движущимися частицами, но и каким-либо возбуждением, распространяющимся со скоростью, превышающей фазовую скорость света в среде. Допустим, например, что на границу раздела двух сред падает волна с плоским фронтом AB (рис. 149). Вдоль границы раздела побежит возмущение со скоростью $V = v_1/\sin\varphi$, где v_1 — фазовая скорость света в первой среде. Оно возбудит во второй среде излучение Вавилова–Черенкова под углом ϑ к границе раздела. Угол ϑ определяется формулой (38.1), т.е. $\cos\vartheta = v_2/V$, где v_2 — фазовая скорость во второй среде. Замечая, что $\cos\vartheta = \sin\psi$, отсюда находим $\sin\psi/\sin\varphi = v_2/v_1$. Это — закон преломления Снеллиуса. Таким образом, преломление света можно трактовать как эффект Вавилова–Черенкова, возбуждаемый во второй среде падающей волной. Так же можно рассматривать и отражение света. Распространение света в однородной среде тоже подчиняется закону (38.1). В этом случае скорость волнового фронта V совпадает с фазовой скоростью v, так что формула (38.1) дает $\vartheta = 0$, т.е. волновой фронт распространяется без изменения направления.

Рис. 149

6. Излучение Вавилова–Черенкова нашло широкое применение в ядерной физике и физике высоких энергий. На нем основано действие так называемых *черенковских счетчиков*, т.е. детекторов релятивистских заряженных частиц, излучение которых регистрируется с помощью фотоумножителей. Основное назначение черенковских счетчиков — разделение релятивистских частиц с одинаковыми импульсами, но различными скоростями. Пусть, например, пучок, состоящий из релятивистских протонов и π-мезонов, проходит через однородное поперечное магнитное поле. Направления траекторий прошедших частиц будут определяться только их импульсами, но не будут зависеть от их скоростей. С помощью диафрагм можно выделить протоны и π-мезоны с одинаковыми импульсами. Из-за различия масс скорости π-мезонов v_π окажутся несколько больше скоростей протонов v_p. Если полученный пучок направить в газ и подобрать показатель преломления n газа так, чтобы было $v_\pi > \frac{c}{n} > v_p$, то π-мезоны будут давать излучение Вавилова–Черенкова, а протоны — нет. Таким образом, счетчик будет регистрировать только π-мезоны, но не будет регистрировать протоны. Показатель преломления газа в камере счетчика n можно менять, изменяя его давление.

7. В заключение остановимся на вопросе, может ли возникнуть электромагнитное излучение, когда заряженная частица движется равномерно в среде с досветовой скоростью, т.е. со скоростью, меньшей фазовой скорости света в рассматриваемой среде. Если среда однородна, то ответ будет отрицательным. Действительно, движущаяся заряженная частица на своем пути, конечно, будет возбуждать атомы и молекулы среды. Последние начнут излучать. Однако эти излучения в результате интерференции погасят друг друга, так как при равномерном движении частицы и однородности среды их амплитуды одинаковы, а фазы возрастают линейно с расстоянием, пройденным частицей. Но если среда неоднородна, то гашения не будет и появится излучение. Такое излучение, на существование которого было указано В.Л. Гинзбургом и И.М. Франком в 1944 г., называется *переходным излучением.*

Особенно наглядно происхождение переходного излучения можно понять на следующем примере. Если электрон (или другая заряженная частица) находится перед плоской границей идеального металла, то электрическое поле вне металла можно рассматривать как поле диполя, состоящего из электрона и его «электрического изображения» в поверхности металла (см. т. III, § 23). Если электрон приближается к металлу, то электрическое изображение движется к нему навстречу. При этом электрический момент диполя уменьшается. Вследствие этого и возникает излучение. В момент, когда электрон пересекает границу металла, происходит как бы аннигиляция электрона и его электрического изображения. Так же возникает переходное излучение, когда электрон выходит из металла в вакуум.

Аналогично объясняется возникновение переходного излучения, когда заряженная частица переходит через границу двух диэлектриков. В этом случае также примени́м метод электрических изображений, хотя и в несколько измененной форме (см. т. III, § 24).

Переходное излучение наблюдалось уже давно в виде свечения анодов рентгеновских трубок. Конечно, природа этого свечения была выяснена много позднее.

Переходное излучение также используется в счетчиках релятивистских частиц для определения их скоростей.

ГЛАВА IV

ДИФРАКЦИЯ СВЕТА

§ 39. Принцип Гюйгенса–Френеля. Зоны Френеля

1. Под *дифракцией света* понимают всякое уклонение от прямолинейного распространения света, если оно не может быть истолковано как результат отражения, преломления или изгибания световых лучей в средах с непрерывно меняющимся показателем преломления. Если в среде имеются мельчайшие частицы постороннего вещества (туман) или показатель преломления заметно меняется на расстояниях порядка длины волны, то в этих случаях говорят о *рассеянии света* и термин «дифракция» не употребляется. Явления дифракции для своего истолкования и количественного рассмотрения не требуют никаких новых принципов. Всякая дифракционная задача, если ее рассматривать строго, сводится к нахождению решения уравнений Максвелла, удовлетворяющего соответствующим граничным условиям. Однако в такой строгой постановке дифракционные задачи, ввиду их сложности, допускают аналитические решения лишь в простейших идеализированных случаях. В оптике значительно большее значение имеют нестрогие методы решения дифракционных задач, основанные на принципе Гюйгенса в обобщенной формулировке Френеля или Кирхгофа.

2. В § 3 мы указывали, что принцип Гюйгенса в том виде, в каком он был сформулирован самим Гюйгенсом, есть не более чем *геометрический рецепт* для построения волновых фронтов. Во всех применениях вторичные волны Гюйгенса выступают не как реальные волны, а как *вспомогательные сферы*, используемые для такого построения. Эти сферы, построенные из точек волнового фронта как из центров, проявляют свое действие *только на огибающей*, которая и дает новое положение волнового фронта. При этом оставалось необъясненным, почему при распространении волны не возникает *обратная волна*.

Искусственную гипотезу об огибающей вторичных волн Френель заменил физически ясным положением, согласно которому вторичные волны при наложении *интерферируют* друг с другом. Свет должен наблюдаться во всех местах пространства, где при интерференции вторичные волны усиливаются; в тех же местах, где они взаимно гасят друг друга, должна наблюдаться темнота. Тем самым выясняется и физический смысл огибающей. К огибающей все вторичные волны приходят в *одинаковых фазах*, и их интерференция приводит к большой интенсивности света. Становится ясным, по крайней мере качественно, и отсутствие обратной волны. Вторичные волны, идущие от волнового фронта вперед, вступают в свободное от возмущений пространство. Они интерферируют только друг с другом. Напротив, вторичные волны, идущие назад, вступают в пространство, где уже есть волновое возмущение — прямая волна. При интерференции вторичные волны *гасят прямую волну*, так что после прохождения волны пространство за ней оказывается невозмущенным.

Френель дал следующую формулировку принципа Гюйгенса, несколько обобщенную Рэлеем (1842–1919). *Окружим все источники света S_1, S_2, S_3, … произвольной замкнутой поверхностью F* (рис. 150). *Каждую точку такой поверхности можно рассматривать как источник вторичных волн, распространяющихся во всех направлениях. Эти волны когерентны, поскольку все они возбуждаются одними и теми же первичными источниками. Световое поле, возникающее в результате их интерференции, в пространстве вне поверхности F совпадает с полем реальных источников света.*

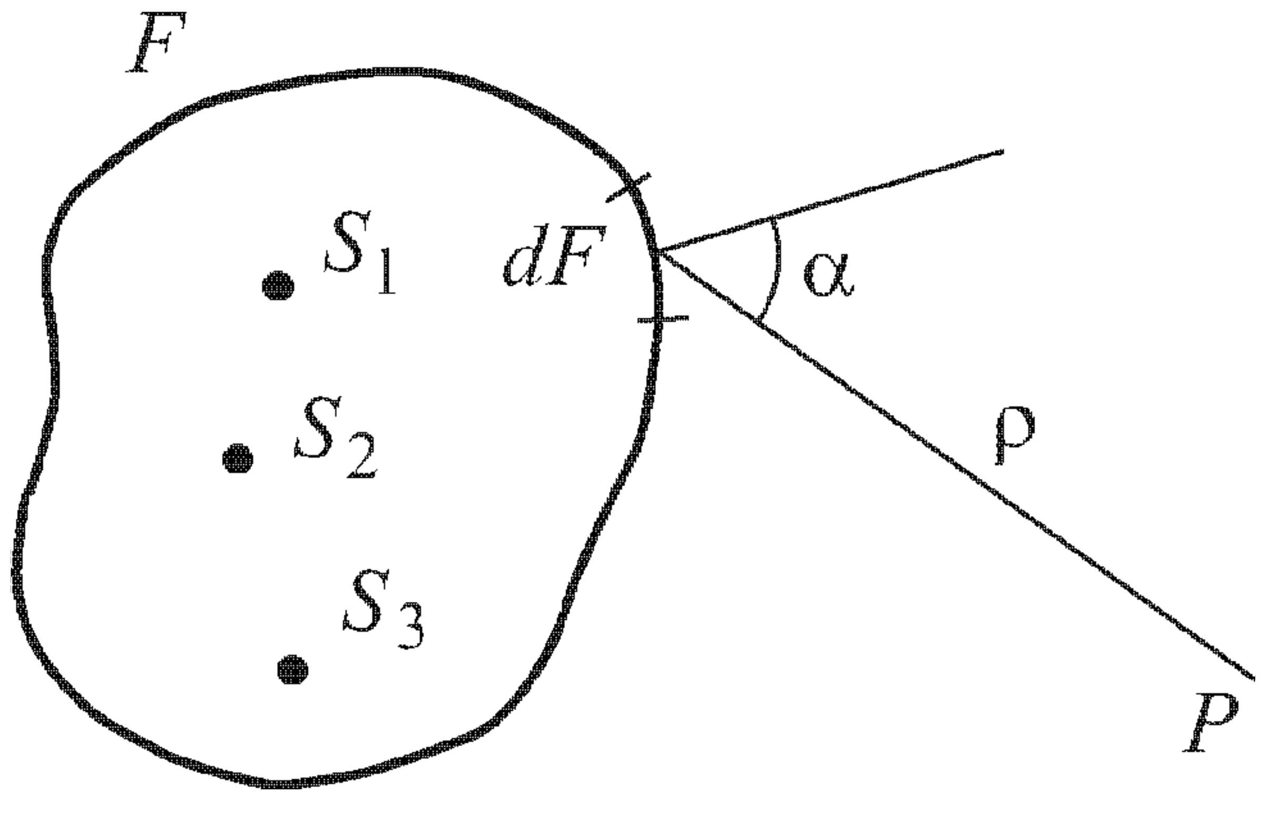


Рис. 150

Таким образом, действительные источники света можно как бы заменить окружающей их светящейся поверхностью F с непрерывно распределенными по ней когерентными вторичными источниками. Отличие этой поверхности от реальной поверхности излучающего тела состоит в том, что она абсолютно прозрачна для всякого излучения. В такой формулировке принцип Гюйгенса–Френеля выражает весьма общее положение. Он означает, что волна, отделившаяся от своих источников, в дальнейшем ведет *автономное существование*, совершенно не зависящее от наличия источников.

3. При математической формулировке принципа Гюйгенса–Френеля будем предполагать, что источники света — монохроматические с одной и той же частотой ω. По предположению Френеля каждый элемент площади dF поверхности F (рис. 150) испускает вторичную сферическую волну, а волновое поле в точке наблюдения P представляется суперпозицией таких волн в виде интеграла

$$E = \int \frac{a\,dF}{\rho} e^{i(\omega t - k\rho)} \tag{39.1}$$

по всей (замкнутой) поверхности F. Френель предположил, что амплитуда вторичной волны пропорциональна амплитуде первичной волны, приходящей к элементу dF, а также площади самого элемента dF. Кроме того, она убывает с возрастанием угла между нормалью к поверхности F и направлением излучения вторичной волны. Конкретного выражения для амплитуды вторичной волны Френель не дал. Это было сделано только в 1883 г. Кирхгофом (1824–1887) при строгой формулировке принципа Гюйгенса–Френеля (см. § 43). Несмотря на это, Френелю удалось получить правильные решения многих дифракционных задач, пользуясь простыми интуитивными соображениями и нестрогими методами расчета. Эти простые и наглядные методы, приводящие быстро к цели, до сих пор используются при элементарном изложении теории дифракции. Они взяты за основу и в настоящей книге. Мы изменим только те (не влияющие на окончательный результат) предположения Френеля, которые оказались неверными.

4. Рассмотрим сначала свободное распространение сферической волны

$$E_0 = \frac{1}{r_0} e^{i(\omega t - kr_0)} \tag{39.2}$$

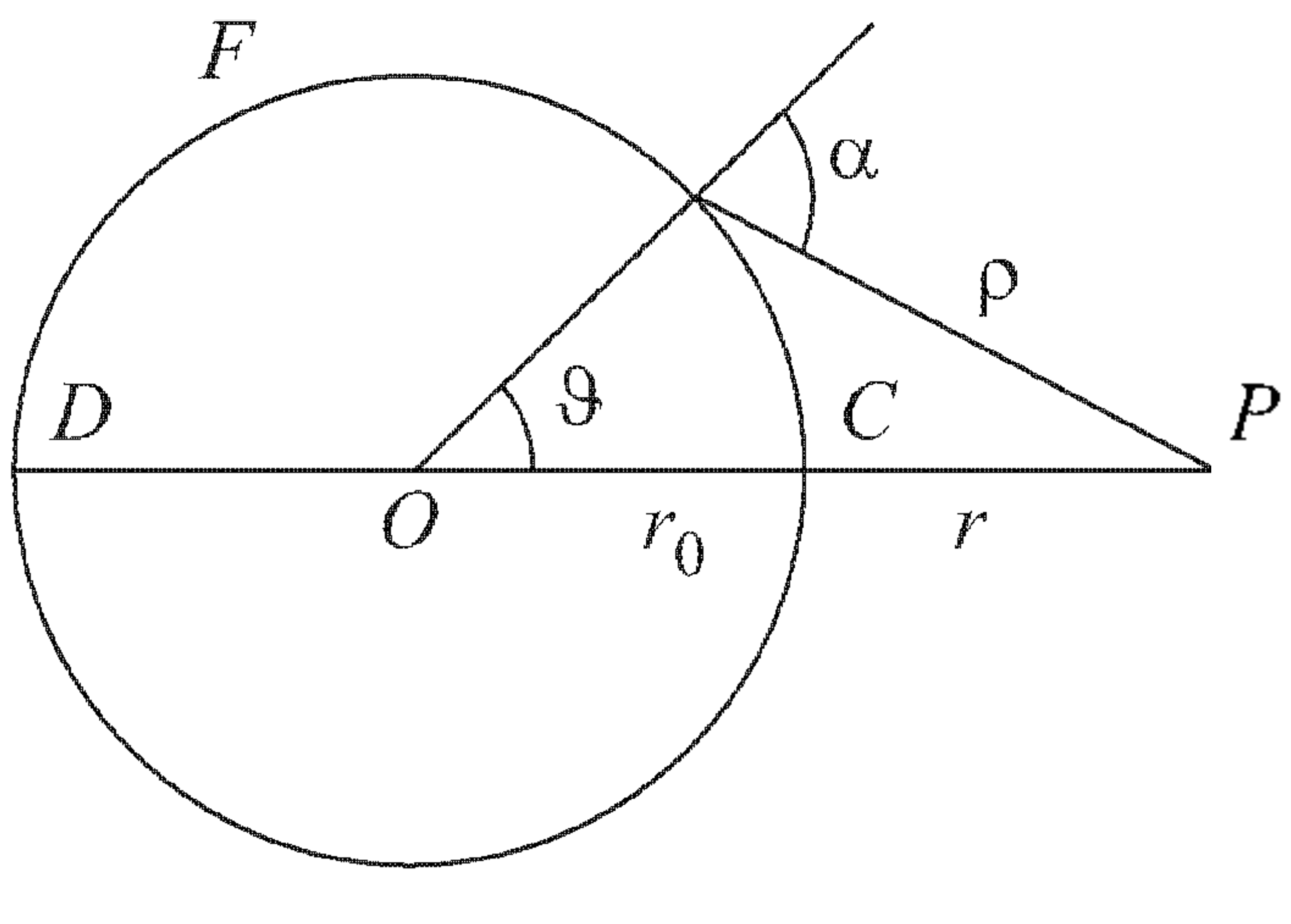


Рис. 151

в однородной среде (рис. 151). Перед этим выражением можно было бы поставить какую-то постоянную амплитуду A_0. Мы не делаем этого, выбирая единицы измерения так, чтобы было $A_0 = 1$. В качестве вспомогательной поверхности F в формуле (39.1) выберем сферический волновой фронт радиуса r_0. Согласно гипотезе Френеля, $a = K(\alpha)e^{-ikr_0}/r_0$, где функция $K(\alpha)$, помимо длины волны, зависит только от угла между нормалью к волновому фронту и направлением излучения вторичной волны Гюйгенса. Полное поле в точке наблюдения P представится интегралом

$$E = \int \frac{K(\alpha)}{r_0 \rho} e^{i(\omega t - kr_0 - k\rho)}\, dF.$$

Если за элемент площади dF принять площадь кольца, вырезаемого из волнового фронта двумя бесконечно близкими концентрическими сферами с центрами в точке наблюдения P, то $dF = 2\pi r_0^2 \sin\vartheta\, d\vartheta$. Примем за переменную интегрирования расстояние ρ. Дифференцируя соотношение $\rho^2 = r_0^2 + (r_0 + r)^2 - 2r_0(r_0 + r)\cos\vartheta$ при постоянных r_0 и r, находим $\sin\vartheta\, d\vartheta$, а затем и элемент площади dF. В результате получим

$$E = \frac{2\pi}{r_0 + r}\, e^{i(\omega t - kr_0)} \int\limits_{r}^{r_{\text{макс}}} K(\rho) e^{-ik\rho}\, d\rho, \tag{39.3}$$

где прежняя функция $K(\alpha)$ теперь рассматривается как функция ρ. Верхний предел интеграла равен $r_{\text{макс}} = r + 2r_0$. Однако, имея в виду дальнейшие применения, мы не будем его конкретизировать. Точное вычисление интеграла (39.3), конечно, невозможно без знания вида функции $K(\rho)$. Однако Френель, используя малость длины световой волны, дал метод приближенного вычисления подобных интегралов при весьма общих предположениях относительно функции $K(\rho)$. Опишем из точки P, как из центра, концентрические сферы с радиусами r, $r + \lambda/2$, $r + 2(\lambda/2)$, $r + 3(\lambda/2)$, ... (рис. 152). Они разобьют волновой фронт F на кольцевые области, получившие название *зон Френеля.* Ввиду малости длины волны, аргумент ρ, а с ним и функция $K(\rho)$ в пределах одной зоны могут считаться постоянными. В этом приближении интеграл (39.3) по n-й зоне будет равен

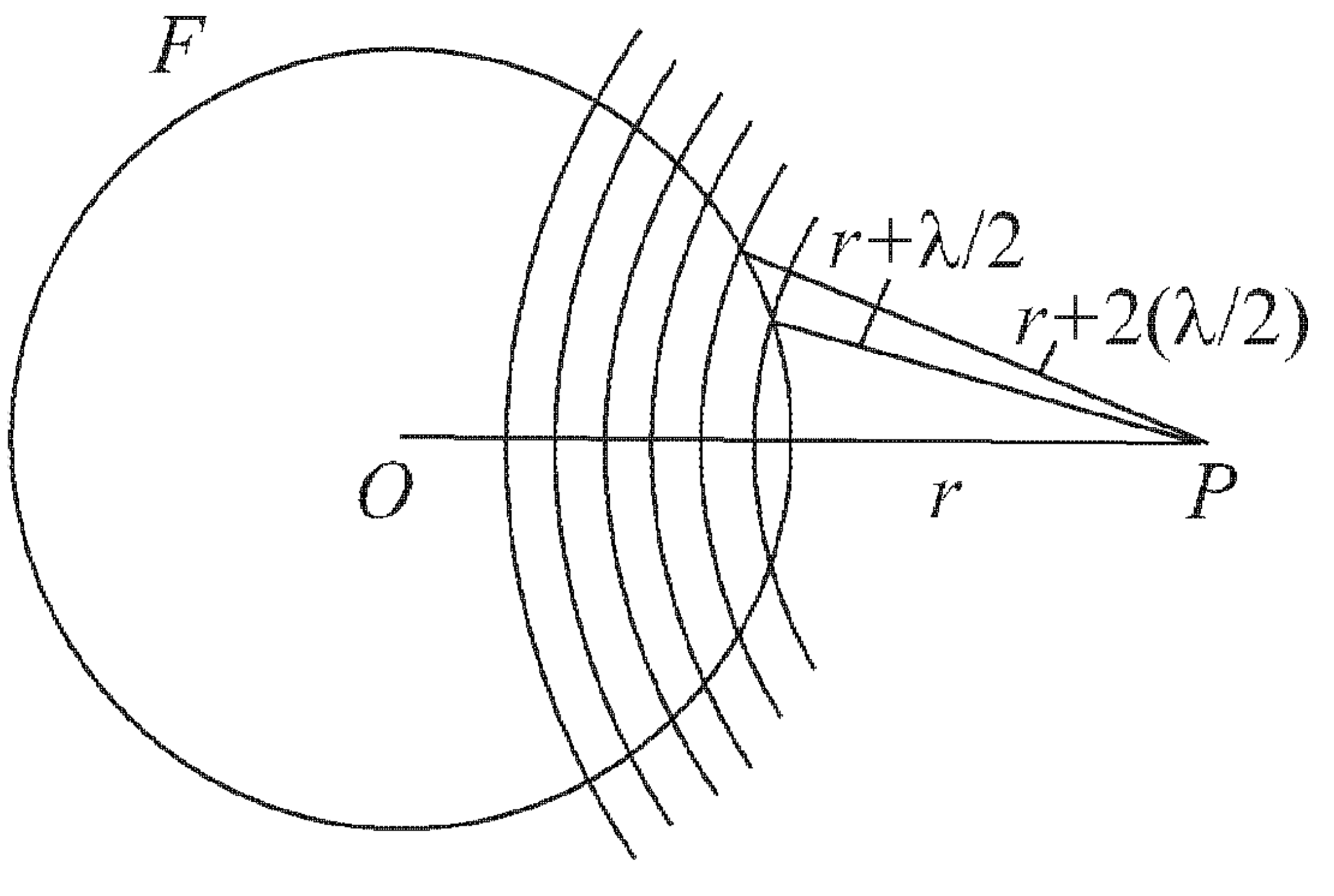


Рис. 152

$$K_n \int\limits_{r + (n-1)\lambda/2}^{r + n\lambda/2} e^{-ik\rho}\, d\rho = (-1)^{n+1} \frac{2K_n}{ik}\, e^{-ikr},$$

где через K_n обозначено среднее значение функции $K(\rho)$ для n-й зоны. Если N — общее число зон, посылающих вторичные волны в точку наблюдения, то поле E представится знакопеременной суммой

$$E = E_1 + E_2 + E_3 + \ldots + E_N. \tag{39.4}$$

Общий член ее имеет вид

$$E_n = (-1)^{n+1} \frac{4\pi K_n}{ik(r_0 + r)}\, e^{i[\omega t - k(r_0 + r)]}. \tag{39.5}$$

То обстоятельство, что знаки рядом стоящих слагаемых в сумме (39.4) противоположны, означает, что колебания, вносимые соседними зонами Френеля, противоположны по фазе. Этого и следовало ожидать, так как из самого построения зон Френеля видно, что одно из этих двух колебаний запаздывает относительно другого на полволны. По предположению Френеля абсолютные значения множителей K_n, а с ними и членов суммы (39.4) медленно убывают с возрастанием номера n. Уберем в сумме (39.4) слагаемое E_1 и добавим слагаемое E_{N+1}. Ввиду медленности убывания членов суммы (39.4), интуитивно можно ожидать, что от этого абсолютное значение суммы (39.4) почти не изменится, а ее знак заменится на противоположный. Приняв это, наряду с (39.4) можно также написать:

$$E = -E_2 - E_3 - \ldots - E_N - E_{N+1}.$$

Тогда $2E = E_1 - E_{N+1}$. Но в том же приближении $E_N = -E_{N-1}$, и следовательно,

$$E = \frac{1}{2}(E_1 + E_n). \tag{39.6}$$

Таким образом, волновое возмущение, создаваемое N первыми зонами Френеля, равно полусумме возмущений, вносимых крайними зонами.

Возьмем теперь все зоны на сферическом волновом фронте F (см. рис. 151). Введем предположение (обосновываемое при строгой формулировке принципа Гюйгенса, данной Кирхгофом), что при $\alpha = \pi$, т.е. в точке D волнового фронта, функция $K(\alpha)$ обращается в нуль. Тогда $E_N = 0$, и из формулы (39.6) получаем

$$E = \frac{1}{2}E_1 = \frac{2\pi K_1}{ik(r_0 + r)}\, e^{i[\omega t - k(r_0 + r)]}, \tag{39.7}$$

т.е. *при свободном распространении волны волновое возмущение от всего волнового фронта составляет половину возмущения, даваемого одной только первой зоной Френеля.* Дело происходит так, как если бы из всего волнового фронта действующей осталась только часть первой зоны Френеля¹).

5. Формула (39.7) интересна также в том отношении, что она позволяет точно определить множитель K_1. Действительно, при свободном распространении сферической волны (39.2)

---

¹) В полученном результате авторы многих книг видят доказательство прямолинейного распространения света, утверждая, что свет распространяется как бы в узком прямолинейном канале, поперечное сечение которого порядка размеров первой френелевой зоны. Однако не надо забывать, что этот вывод относится к *свободному* распространению волны, когда проблема прямолинейного распространения света вообще не возникает. Она возникает лишь *при наличии препятствий* на пути распространения света. Это есть *проблема теней*, являющаяся частным случаем *проблемы дифракции света*.

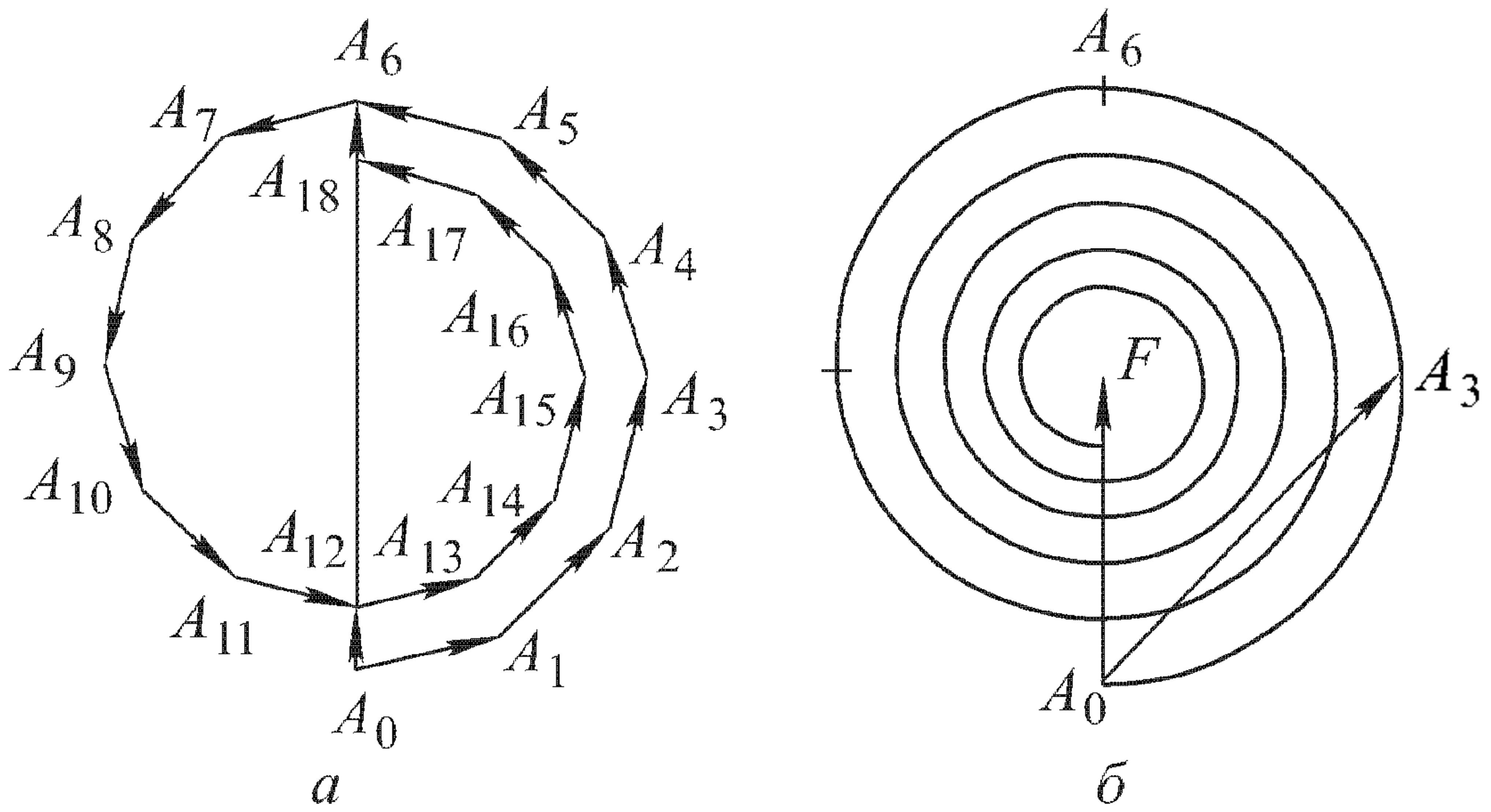
напряженность поля заранее известна в любой точке пространства. В частности, в точке наблюдения

$$E = \frac{1}{r_0 + r}\, e^{i[\omega t - k(r_0 + r)]}.$$

Сравнивая это выражение с (39.7), находим

$$K_1 = \frac{ik}{2\pi} = \frac{k}{2\pi}\, e^{i\pi/2}. \tag{39.8}$$

Таким образом, множитель K_1, а также все остальные множители K_2, K_3, ..., — чисто мнимые. Значит, *вторичные волны Гюйгенса опережают по фазе колебания поля* в точках волнового фронта на $\pi/2$. Если бы такого опережения не было, то вторичные волны из центра первой френелевой зоны приходили бы в точку наблюдения с фазой $\omega t - k(r_0 + r)$, а от ее краев — с меньшей фазой $\omega t - k(r_0 + r) - \pi$. Фаза результирующего колебания, возбуждаемого всей первой зоной, была бы равна полусумме этих значений, т.е. оказалась бы меньше правильного значения на $\pi/2$.



а *б*

Рис. 153

6. Результаты (39.6) и (39.7) имеют основное значение в методе зон Френеля. Для лучшего уяснения интерпретируем их на векторной диаграмме (см. т. III, § 126). Разобьем каждую зону Френеля на m кольцевых подзон (на рис. 153 *а* построение выполнено для $m = 6$). Колебания, возбуждаемые в точке наблюдения такими подзонами, на векторной диаграмме изобразятся векторами $\overrightarrow{A_0A_1}$, $\overrightarrow{A_1A_2}$, $\overrightarrow{A_2A_3}$ и т.д., образующими ломаную линию. Колебание, возбуждаемое несколькими соседними подзонами, представится геометрической суммой таких векторов. Например, вектор $\overrightarrow{A_0A_6}$ представляет «действие» первой френелевой зоны, вектор $\overrightarrow{A_0A_{12}}$ — совместное действие двух первых френелевых зон, и т.п. Если число подзон N устремить к бесконечности, то в пределе ломаная перейдет в непрерывную спираль, вьющуюся вокруг фокуса F (рис. 153 *б*). На этой спирали действие первой френелевой зоны представляется вектором $\overrightarrow{A_0A_6}$, а всего волнового фронта — вектором $\overrightarrow{A_0F}$, который почти точно вдвое короче предыдущего. Действие центральной половины первой зоны представляется наклонным вектором $\overrightarrow{A_0A_3}$, и т.п.

7. В § 27 (п. 8) указывалось, что прохождение сходящейся сферической волны через фокус (т.е. ее центр) сопровождается изменением фазы на π. Приведем новое истолкование этого явления, используя метод зон Френеля.

Как всегда, зоны Френеля надо строить с той стороны волнового фронта, *куда он распространяется*. В рассматриваемом случае это будет вогнутая сторона. Конечно, и здесь вторичные волны будут опережать по фазе колебания поля на волновом фронте, так как это есть свойство *самих элементарных источников вторичных волн*, не зависящее от формы волнового фронта, на котором они находятся. Действие же всего волнового фронта в точке наблюдения при свободном распространении волны по-прежнему будет равно половине действия первой френелевой зоны. В частности, вторичные волны, исходящие от этой зоны, при интерференции всегда дают правильную фазу, независимо от положения точки наблюдения. Если точка наблюдения P лежит перед центром (фокусом) O сходящейся сферической волны (рис. 154), то в наших предыдущих рассуждениях ничего не изменится, так как в этом случае колебания от краев A и B первой френелевой зоны приходят в P позже, чем от ее середины C. Но положение изменится, когда точка наблюдения P' будет находиться по другую сторону от центра O (тогда, конечно, и первая зона $A'B'$ будет другой). Теперь края первой зоны A' и B' окажутся ближе на полволны к точке P', чем ее центр C. Вследствие этого, помимо обычного опережения по фазе на $\pi/2$, вторичные волны, приходящие в P' от краев A' и B' первой зоны, получат дополнительное приращение фазы на π, так что полное опережение по фазе вследствие указанных двух причин составит $\frac{1}{2}\left(\frac{\pi}{2}+\frac{3\pi}{2}\right)=\pi$. Его надо добавить к фазе волны $\omega t - kr$, прошедшей через точку O. В результате мы и приходим к явлению изменения фазы сходящейся сферической волны при ее прохождении через фокус.

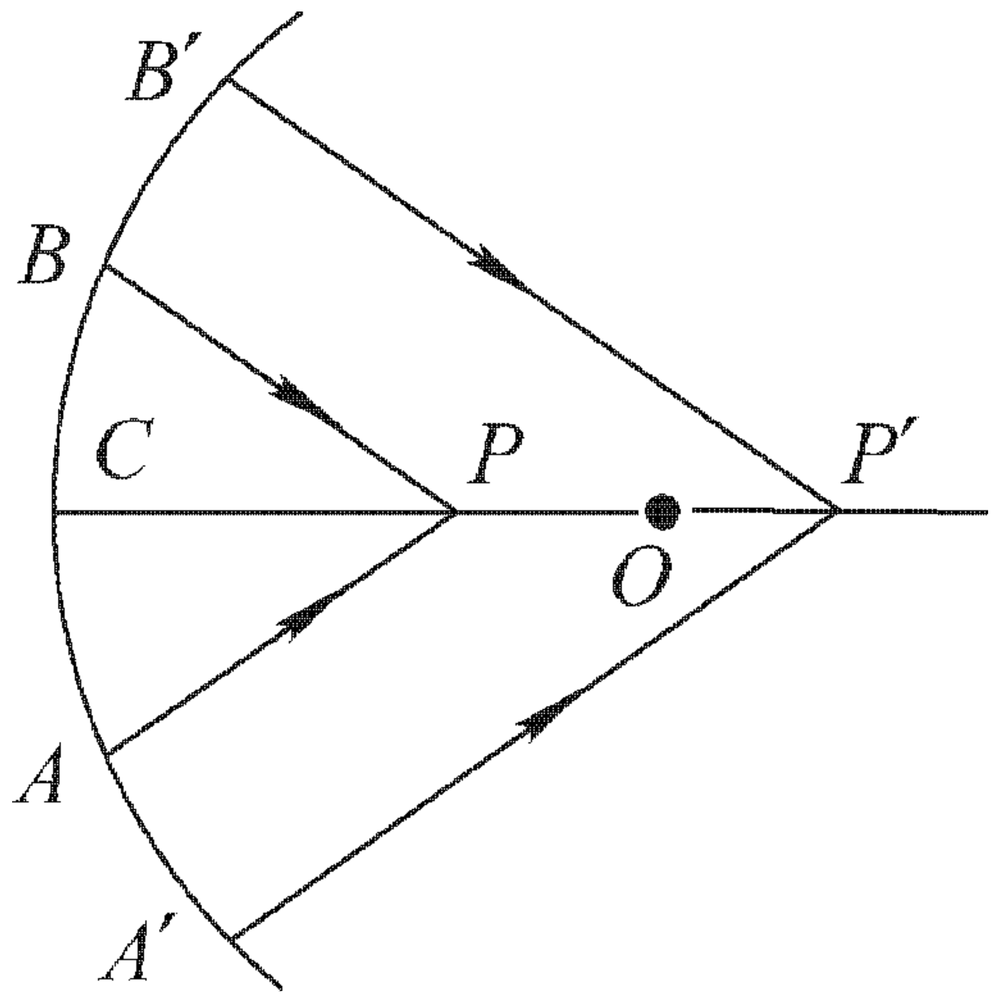


Рис. 154

Когда точка наблюдения P приближается к фокусу O, френелевы зоны расширяются. В некотором положении P_0 этой точки первая френелева зона охватывает всю сферу. При дальнейшем приближении к O построение зон Френеля становится невозможным. Можно сказать, что вся сфера становится как бы частью первой зоны Френеля. Применяя к этому случаю наши предыдущие рассуждения, нетрудно убедиться, что появляется опережение по фазе, заключенное между нулем и $\pi/2$, когда точка наблюдения находится между P_0 и O. В точке O оно составляет $\pi/2$, а при переходе через нее возрастает до π. Таким образом, изменение фазы волны на π при переходе через фокус совершается не скачком, а непрерывно.

§ 40. Дифракция на оси от круглого отверстия и экрана. Зонная пластинка

1. Поставим между точечным источником S и точкой наблюдения P непрозрачный экран с круглым отверстием, плоскость которого перпендикулярна к оси SP, а центр O расположен на той же оси (рис. 155). Согласно Френелю, действие такого препятствия сводится к тому, что экран как бы устраняет ту часть волнового фронта, которую он прикрывает. На открытой же части волнового фронта световое поле не изменяется. Такое предположение соответствует приближению геометрической оптики, а потому оно может быть приближенно верным только тогда, когда радиус отверстия очень велик по сравнению с длиной световой волны. Это и предполагается в дальнейшем. Будем предполагать, кроме того, что размеры отверстия можно менять, что дает возможность открывать любое число зон Френеля.

Рис. 155

Пусть a_0 и I_0 означают амплитуду и интенсивность света в точке P при свободном распространении волны, т.е. в отсутствие экрана. Общий световой поток, поступающий через отверстие, строго пропорционален его площади. Но он будет распределяться по-разному по освещаемой поверхности, в зависимости от того, сколько, зон Френеля укладывается в отверстии. В одних местах может получиться интенсивность меньше, в других больше I_0. Никакого противоречия с законом сохранения энергии в этом нет.

Если открыть первую зону Френеля, то амплитуда и интенсивность света в той же точке будут $a_1 = 2a_0$, $I_1 = (2a_0)^2 = 4I_0$. Таким образом, интенсивность в центре P дифракционной картины получится в четыре раза большей, чем было бы при свободном распространении волны. При удалении от центра P интенсивность будет монотонно убывать. При расширении отверстия в точку P начнут приходить вторичные волны. Их интерференция с ранее пришедшими волнами вызовет уменьшение интенсивности в той же точке. При определенных размерах отверстия центр P перестанет быть точкой максимальной интенсивности. Вокруг точки P начнет образовываться светлое кольцо, к которому и переместится максимум интенсивности. Когда отверстие откроет две первые зоны Френеля, то их действия в точке P практически полностью уничтожат друг друга из-за интерференции. В точке P получится темный кружок, окруженный светлым кольцом.

При дальнейшем увеличении размеров отверстия действия двух первых зон Френеля в точке P остаются компенсированными. Все поле в P создается только частью третьей зоны Френеля. В центре картины появляется светлое пятнышко, а центральный темный кружок расширяется и переходит в темное кольцо, окружающее это пятнышко. Когда число N открытых зон равно трем, интенсивность света в точке P будет такой же, какая получилась бы, если бы была открыта только одна третья зона. Центр картины будет практически столь же светлым, что и при одной открытой первой зоне Френеля. При $N = 4$ светлый центр сменится темным. Вообще, при нечетном N центр дифракционных колец светлый, а при четном — темный. На схематическом рис. 156 для различных N показано распределение интенсивности света в зависимости от расстояния до центра картины, а рис. 157 воспроизводит фотографии наблюдаемой картины колец при нечетном (*а*) и четном (*б*) N.

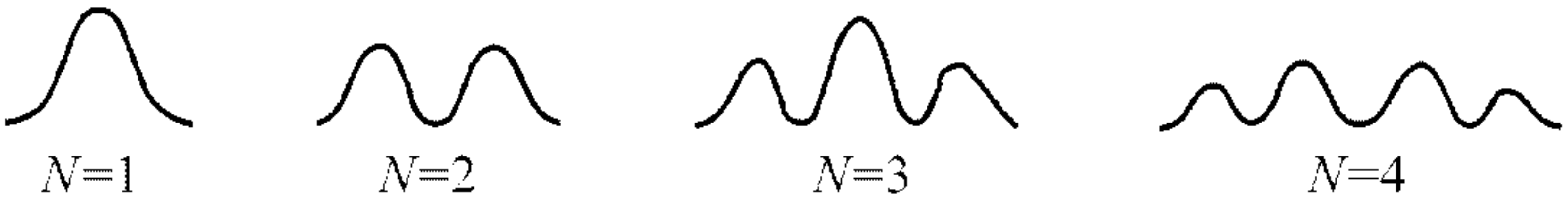


Рис. 156

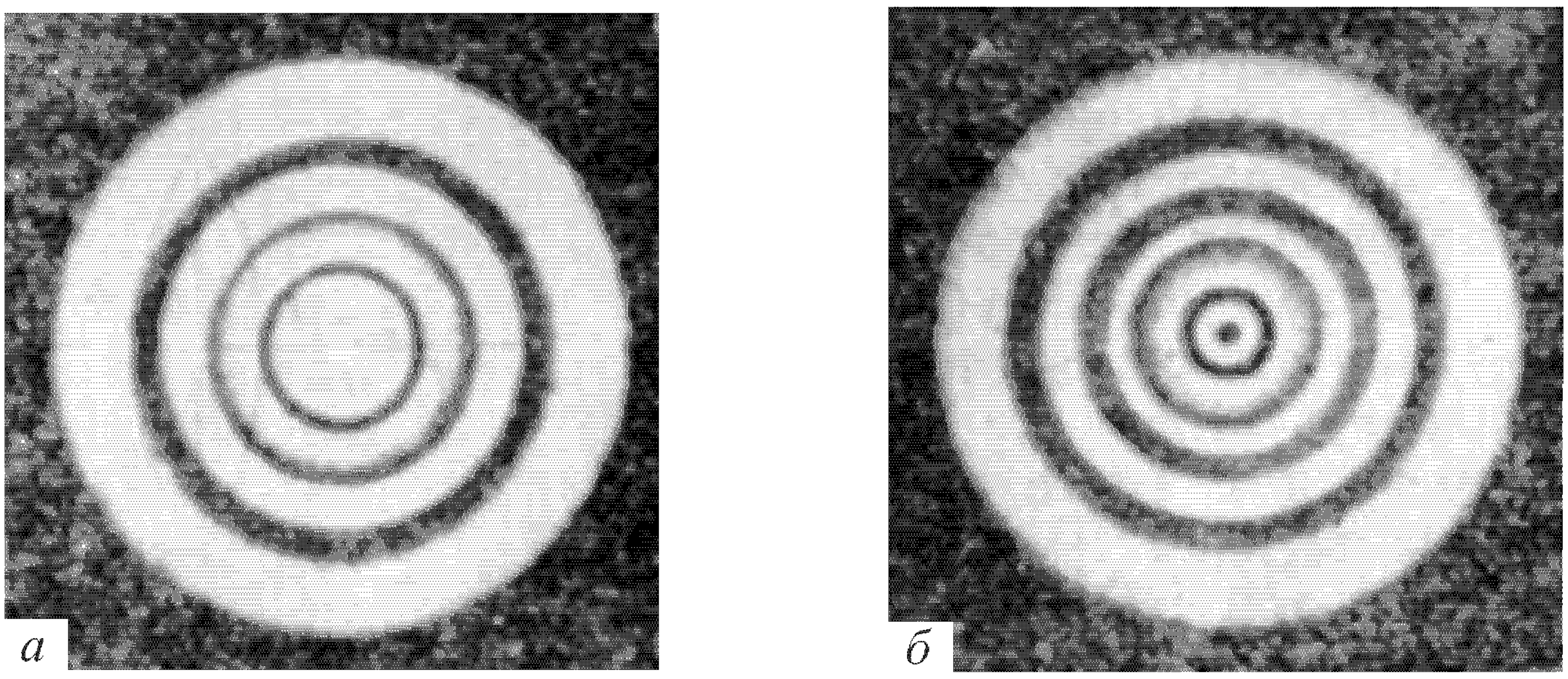


Рис. 157

2. Определим теперь размеры и число m зон Френеля, укладывающихся в отверстии AB (см. рис. 155). Пусть D — диаметр отверстия, а a и b — расстояния от его центра до точек S и P. Из точек S и P как из центров опишем сферы, проходящие через край отверстия AB. Пренебрегая квадратами отрезков OE и OF, по известной геометрической теореме можем написать:

$$(D/2)^2 \approx OF \cdot 2a, \quad (D/2)^2 = OE \cdot 2b.$$

Отсюда

$$EF = EO + OF = \frac{D^2}{8}\left(\frac{1}{a} + \frac{1}{b}\right).$$

Число m найдется делением этого отрезка на $\lambda/2$. Оно равно

$$m = \frac{D^2}{4\lambda}\left(\frac{1}{a} + \frac{1}{b}\right). \tag{40.1}$$

Если m целое, то D будет диаметром, а $R_m = D/2$ — радиусом m-й зоны, точнее — ее внешнего края. Следовательно,

$$R_m = \sqrt{\frac{ab}{a+b}\, m\lambda}. \tag{40.2}$$

Например, если $a = b = 1$ м, $\lambda = 600$ нм, то $R_1 \approx 0{,}55$, $R_2 \approx 0{,}77$, $R_3 \approx 0{,}95$ мм и т.д.

3. Интенсивность света в точке наблюдения P можно во много раз усилить, прикрыв все четные или все нечетные зоны Френеля. Оставшиеся неприкрытыми зоны будут усиливать действие друг друга. Прикрытие можно осуществить, поместив в плоскости отверстия так называемую *зонную пластинку* (рис. 158). Ее можно изготовить, начертив на листе бумаги темные кольца, а затем сфотографировав их в уменьшенном масштабе. Внутренние радиусы колец должны быть пропорциональны квадратным корням из последовательных нечетных чисел, а внешние — из четных. Тогда получится пластинка, центр которой светлый. Можно изготовить аналогичную пластинку с темным центром. Ширина всех колец должна быть велика по сравнению с длиной волны. Тогда при надлежащих размерах колец пластинка со светлым центром будет удалять из волнового фронта все четные, а пластинка с темным центром — все нечетные зоны Френеля.

Рис. 158

Усиление интенсивности света зонной пластинкой аналогично фокусирующему действию линзы. Более того, расстояния от пластинки до источника S и «изображения» P связаны тем же соотношением, что и соответствующие величины для линзы. Это видно, если формулу (40.1) переписать в виде

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{f}, \tag{40.3}$$

где «фокусное расстояние» определяется формулой

$$\frac{1}{f}=\frac{4m\lambda}{D^2}=\frac{m\lambda}{R_m^2}. \tag{40.4}$$

Если центр зонной пластинки светлый, то число m — нечетное, в этом случае в формулу (40.4) входит (внешний) радиус светлого кольца пластинки. Если же центр пластинки темный, то число m — четное и под R_m следует понимать (внешний) радиус темного кольца. Какой номер брать при вычислении f — это, конечно, не имеет значения.

С помощью зонной пластинки можно даже получать оптические изображения, хотя и весьма низкого качества.

В отличие от линзы, зонная пластинка имеет *несколько фокусов*. Действительно, найдем положение точки наблюдения P, при котором в центральном круге пластинки уместятся первые три зоны Френеля. Тогда в следующее кольцо пластинки попадут четвертая, пятая и шестая зоны и т.д. Если центр зонной пластинки светлый, то поле в точке P представится суммой

$$E=(E_1+E_2+E_3)+(E_7+E_8+E_9)+(E_{13}+E_{14}+E_{15})+\ldots,$$

или $E=E_1+E_7+E_{13}+\ldots$, так как действия соседних зон практически уничтожают друг друга. Таким образом в P получится максимум (фокус). Фокусное расстояние найдется по формуле $f_1=R_1/(3\lambda)=f/3$. Аналогично находятся фокусы высших порядков:

$$f_n=\frac{f}{2n+1}, \tag{40.5}$$

где n — целые числа, которым можно придавать не только положительные, но и отрицательные значения. Отрицательным значениям соответствуют расходящиеся волны и *мнимые фокусы.*

Рэлей указал, что интенсивность света в точке наблюдения P увеличится в четыре раза, если изменить на π фазы вторичных волн, исходящих от всех зон Френеля с четными (или нечетными) номерами. Такая «*зонная пластинка с обращением фазы*» была изготовлена Вудом (1868–1955) путем травления поверхности стеклянной пластинки. Ее действие вполне эквивалентно действию линзы, поскольку в обоих случаях вторичные волны от всех точек волнового фронта приходят в P в одинаковых фазах.

4. Поставим теперь между источником S и точкой наблюдения P непрозрачный круглый экран AB (рис. 159), плоскость которого перпендикулярна к оси SP. Пусть DA и BE — неприкрытые части волнового фронта сферической волны, исходящей из источника S. Разобьем ее на кольцевые зоны Френеля, начав их построение от края экрана. Рассуждая как и раньше, представим напряженность поля излучения в точке P в виде половины напряженности, создаваемой в этой точке вторичными волнами первой кольцевой зоны Френеля. Следовательно, каков бы ни был диаметр диска, в центре P его геометрической тени должно наблюдаться *светлое пятнышко.* Такой вывод был сделан Пуассоном (1781–1840) и показался последнему столь абсурдным, что он выдвинул его в качестве возражения против волновой теории света Френеля. Араго (1786–1853) немедленно поставил опыт и обнаружил пятнышко в соответствии с выводом Пуассона[1]. Явление получило название *пятна Араго–Пуассона.* Наблюдаемая картина дифракции от круглого экрана приведена на рис. 160.

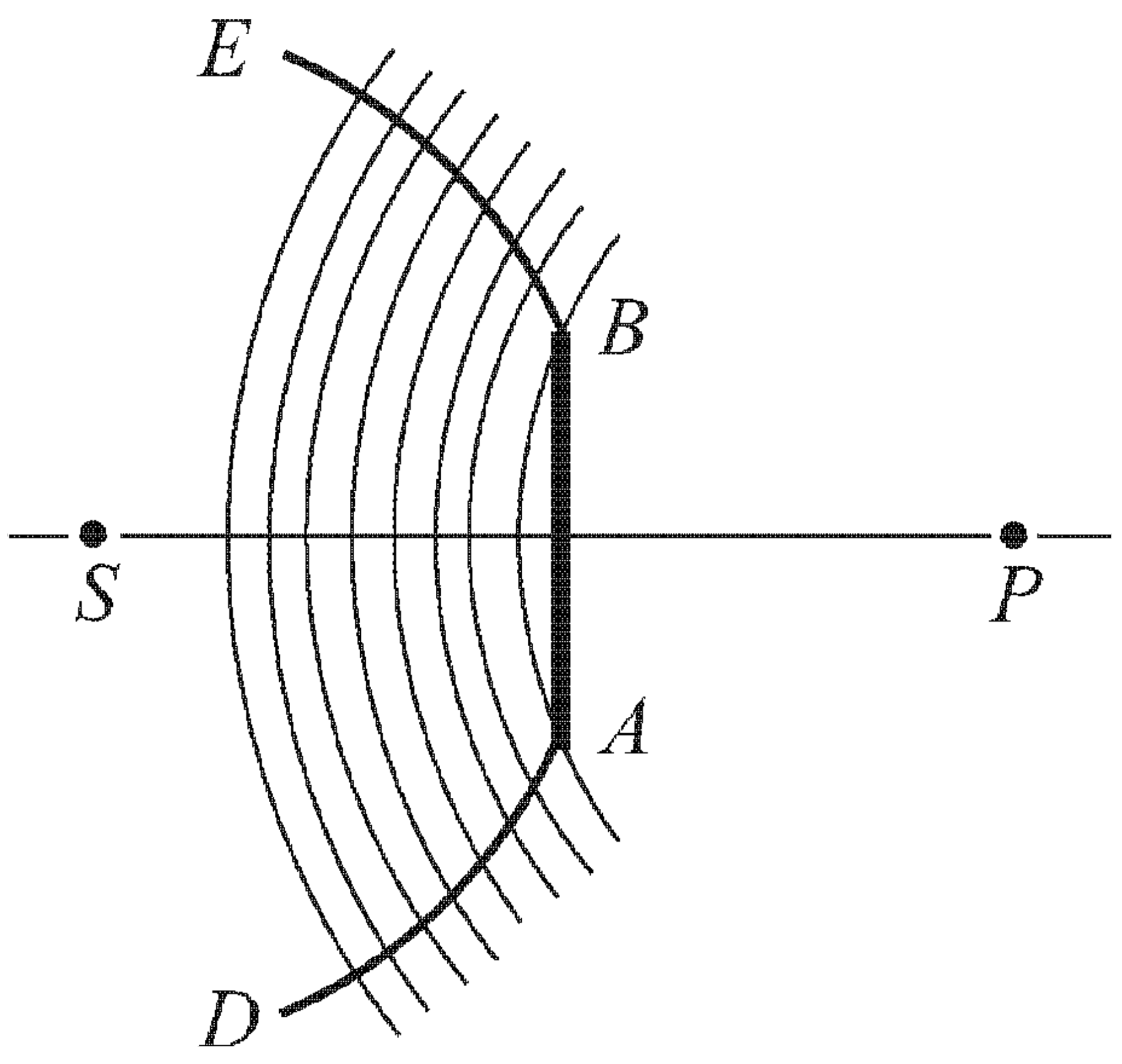


Рис. 159

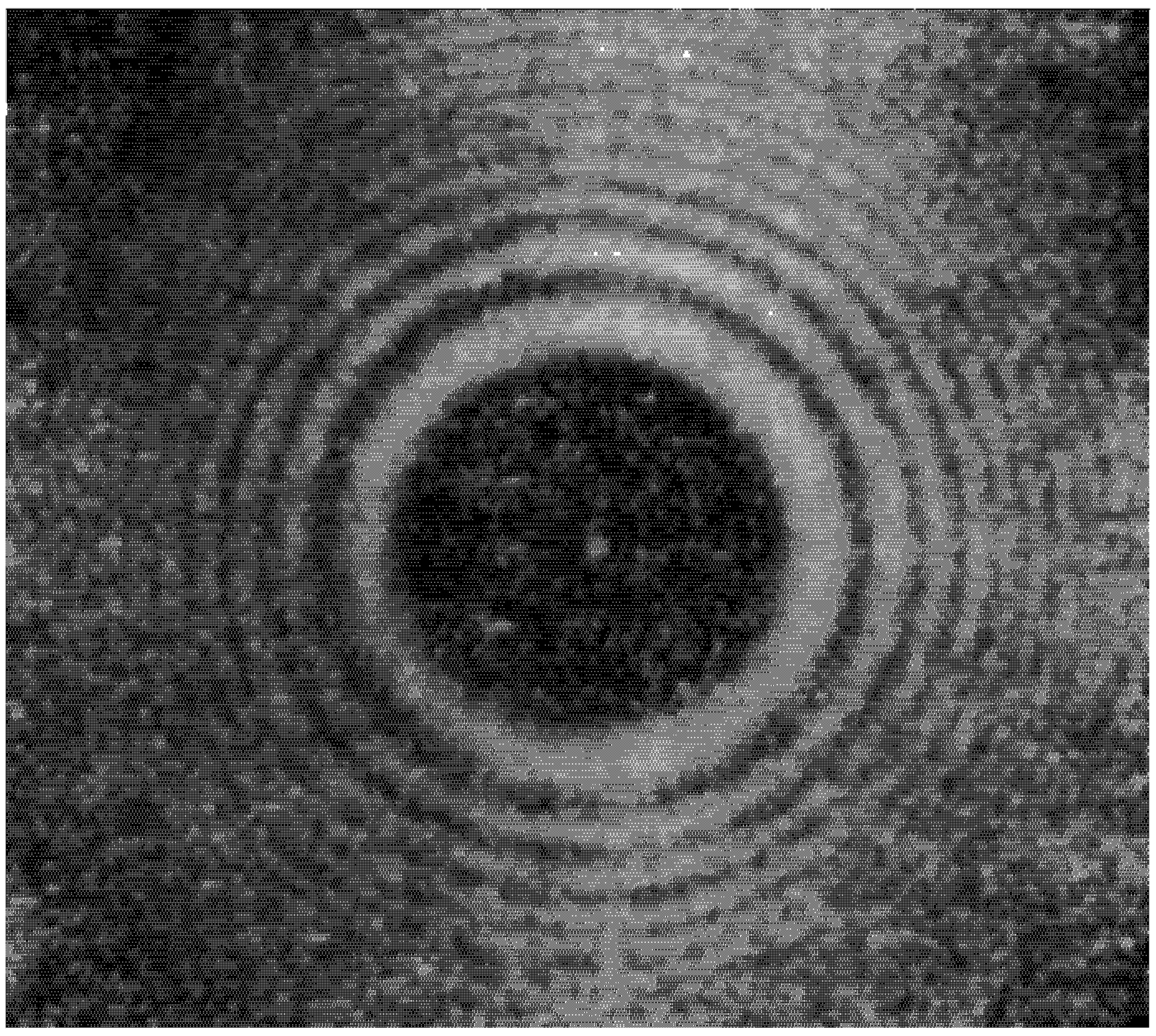


Рис. 160

Если точка наблюдения P не находится в центре картины, то кольцевые зоны Френеля, конечно, можно построить и для нее. Однако зоны, расположенные ближе к центру, окажутся неприкрытыми лишь частично. Это сильно усложняет вычисление интенсивности света. Можно только сказать, что дифракционная картина должна обладать осевой симметрией. Вне геометрической тени получается система концентрических светлых и темных колец. Внутри самой геометрической тени также могут получиться дифракционные кольца, в особенности когда экран прикрывает небольшое число зон Френеля. Но эти кольца мало контрастны, а распределение света в них сложное.

---

[1] Маральди наблюдал такое пятнышко еще в 1723 г. Вероятно, еще раньше (1715 г.) оно наблюдалось Делилем, хотя указания последнего недостаточно ясны. Однако эти наблюдения остались незамеченными и были забыты, поскольку природа явления не была понята.

Демонстрацией явления пятна Араго–Пуассона может служить опыт Поля. Поль сфотографировал ярко освещенный шаблон, заменив объектив гладким металлическим шаром. Эта фотография воспроизведена на рис. 161, а параметры установки приведены в задаче в конце этого параграфа. Ангерер сфотографировал изображение человеческого лица, заменив шар металлическим диском.

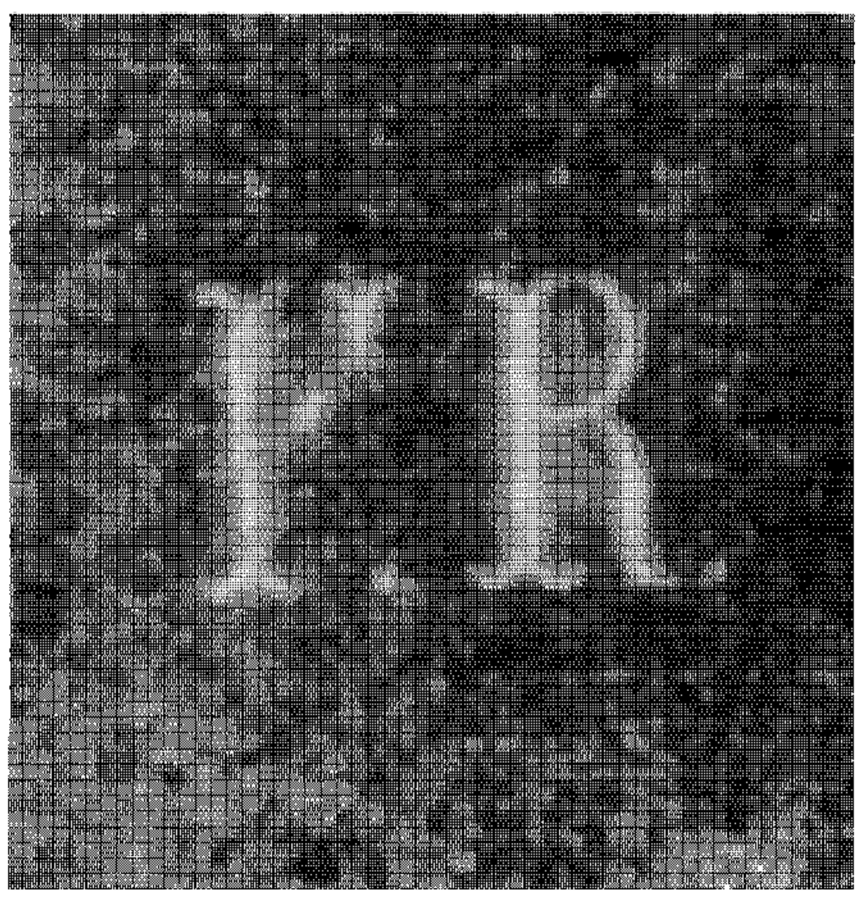


Рис. 161

5. Применим метод зон Френеля к объяснению *теней*, т.е. *прямолинейного распространения света*. Поскольку речь идет о законе, от которого принципиально должны наблюдаться отступления, наши рассуждения не могут претендовать на строгость.

Пусть на пути распространяющейся волны поставлен экран или отверстие произвольной формы. Их размеры должны быть велики по сравнению с длиной волны. Разобьем волновой фронт на кольцевые зоны Френеля. Некоторые зоны могут оказаться открытыми полностью, другие частично, третьи совсем закрытыми.

Допустим сначала, что точка наблюдения лежит вне геометрической тени, далеко от ее границы. Первые члены ряда (39.4) получатся такими же, как и при свободном распространении волны. Последующие члены начнут изменяться из-за частичного экранирования соответствующих им зон. В зависимости от формы края экрана эти изменения будут носить более или менее нерегулярный характер. Если обнаружится тенденция убывания их по абсолютной величине, то правдоподобно допустить, что напряженность поля в точке наблюдения окажется равной половине напряженности, создаваемой центральной зоной. В случае точечного источника и ровного края экрана это может оказаться и не совсем так. Однако если источник не совсем точечный, а края экрана не совсем ровные, то произойдет *статистическое сглаживание* при наложении дифракционных картин от точечных источников, на которые можно разложить протяженный источник. Тогда вдали от края экрана получится такая же освещенность, какая получилась бы при свободном распространении волны.

Допустим теперь, что точка наблюдения лежит внутри геометрической тени, опять далеко от ее границы. Первые зоны Френеля будут полностью закрыты. Нумерацию зон начнем с первой (частично) открытой зоны. Представим поле рядом

$$E = (E_1 + E_2 + \ldots + E_N) + (E_{N+1} + E_{N+2} + \ldots),$$

в котором $N+1$ означает номер первой целиком открытой зоны. Ко второй скобке применимы рассуждения, применявшиеся выше в случае свободного распространения волны. Поэтому вторая скобка приближенно равна $^1\!/_2 E_{N+1}$. В первой скобке слагаемые меняются более или менее нерегулярно, обнаруживая в среднем тенденцию возрастания по абсолютной величине. При статистическом усреднении (с учетом неполной точечности источника и неровностей краев экрана) эти нерегулярности сглаживаются, так что первую скобку можно принять равной $^1\!/_2(E_1 + E_N)$, или $^1\!/_2 E_N$, поскольку величина E_1 должна считаться близкой к нулю из-за малости открытой части соответствующей зоны. Таким образом, $E = {}^1\!/_2(E_N + E_{N+1}) \approx 0$, так как для соседних зон $E_N \approx -E_{N+1}$. Итак, при погружении в область геометрической тени интенсивность света обращается в нуль.

Если источник точечный, а края экрана резкие, то граница геометрической тени расщепляется в дифракционные полосы, как это мы видели при рассмотрении дифракции на круглых отверстии и экране. Однако если края экрана неровные, то полосы начинают размываться, а при увеличении размеров источника переходят в полутень.

ЗАДАЧИ

1. В опыте Поля, описанном в тексте, диаметр шара $D = 40$ мм, расстояние от фотографируемого шаблона до шара $a = 12$ м, расстояние от шара до изображения $b = 18$ м, размер шаблона $y = 7$ мм. Определить размер его изображения y'. При каких условиях опыт удастся с шаром, поверхность которого испещрена множеством неправильных царапин?

Ответ. $y' = \dfrac{b}{a} y = 10{,}5$ мм. Для удачи опыта необходимо, чтобы глубина царапин h не превосходила ширины крайней френелевой зоны:

$$h = \frac{\lambda}{D}\frac{ab}{a+b} = 180\lambda = 0{,}1 \text{ мм}.$$

2. Оценить максимальные угловые размеры α предмета, который можно сфотографировать с помощью непрозрачного диска с идеально гладкими краями.

Решение. При смещении точечного объекта в сторону с главной оптической оси диск представится эллипсом. Малая полуось эллипса будет отличаться от радиуса диска на $r\left(1 - \cos\dfrac{\alpha}{2}\right)$. Это отличие не должно превышать ширины крайней френелевой зоны. Из этого условия для максимально допустимых угловых размеров предмета находим

$$\alpha < \frac{4}{D}\sqrt{\frac{\lambda ab}{a+b}} \approx 0{,}2 \text{ рад} \approx 10^\circ,$$

где использованы обозначения и численные данные из предыдущей задачи.

3. Зонная пластинка применяется для фотографирования предмета, который виден из места нахождения пластинки под углом $\alpha = 0{,}1$ рад. Оценить оптимальное число зон пластинки для получения наибольшей яркости и отчетливости изображения.

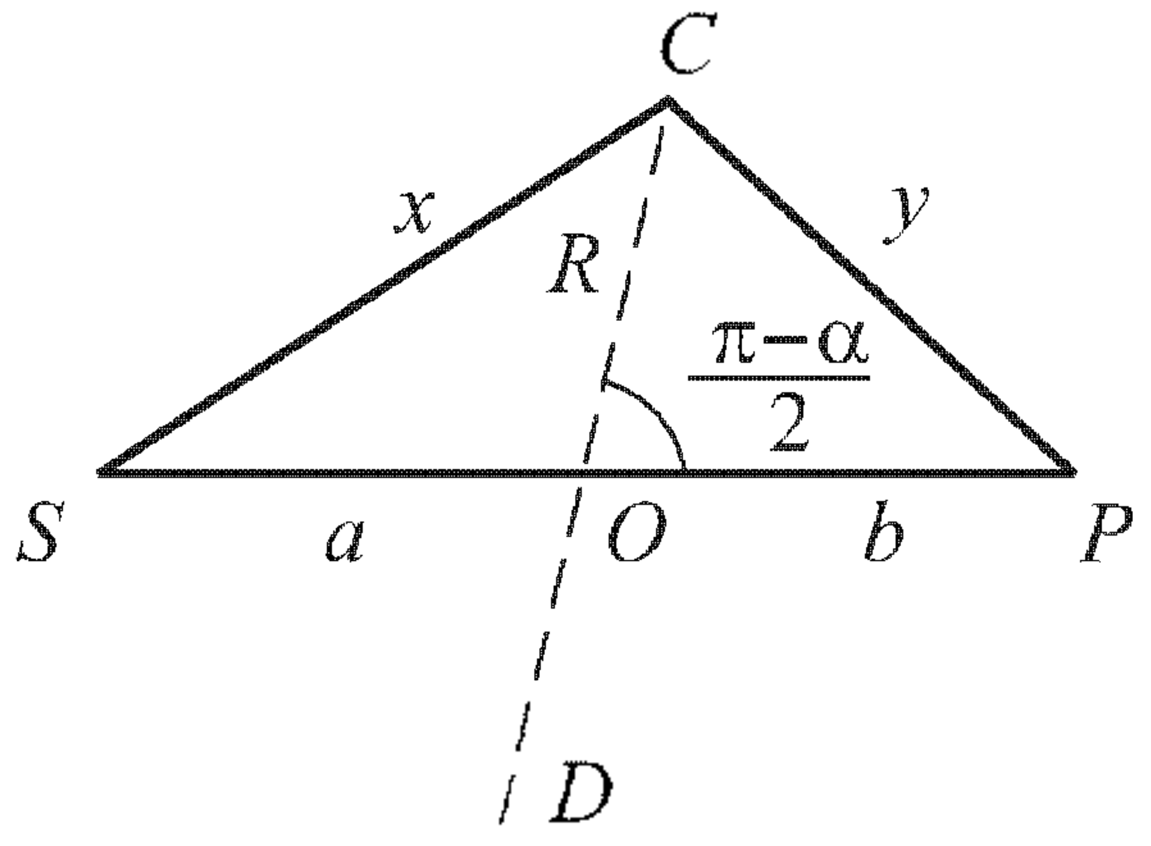


Рис. 162

Решение. Допустим сначала, что источник света S точечный, а зонная пластинка CD наклонена к оптической оси под углом $(\pi-\alpha)/2$ (рис. 162). Из рисунка видно: $x^2 = a^2 + R^2 + 2aR\sin\frac{\alpha}{2}$, и аналогично для y. Извлекая квадратные корни и пренебрегая всеми степенями радиуса R, начиная с третьей, получим для разности хода между лучами SCP и SOP:

$$\Delta = (x+y) - (a+b) = \frac{R\cos^2(\alpha/2)}{2}\left(\frac{1}{a}+\frac{1}{b}\right),$$

или $\Delta = \Delta_0 - \delta\Delta$. Здесь Δ_0 — значение Δ при отсутствии наклона зонной пластинки, а $\delta\Delta$ — приращение величины Δ, обусловленное наклоном:

$$\delta\Delta = \Delta_0\left(1-\cos^2\frac{\alpha}{2}\right) = \Delta_0\sin^2\frac{\alpha}{2} \approx \frac{1}{4}\Delta_0\alpha^2.$$

Если $\delta\Delta \ll \lambda$, то наклон пластинки не скажется существенно на работе зон, расположенных в пределах круга радиуса R. Если же $\delta\Delta \approx \lambda/2$, то все зоны, расположенные выше этого круга, становятся бесполезными и даже вредными. Из этого условия находится предельное значение разности хода: $\Delta_0 \approx 2\lambda/\alpha^2$. Соответствующее число зон Френеля будет

$$N \approx \frac{\Delta_0}{\lambda/2} \approx \frac{4}{\alpha^2}.$$

Допустим теперь, что фотографируемый предмет не точечный, причем его центр расположен на оси зонной пластинки. Для периферийных точек предмета, не лежащих на оси пластинки, последняя действует как наклоненная под углом $\alpha/2$. Поэтому предельное число зон Френеля, при котором должно получиться наиболее отчетливое изображение, будет

$$N \sim 4/\alpha^2 \sim 400.$$

§ 41. Метод Френеля решения дифракционных задач. Дифракция Фраунгофера и Френеля

1. Принципа Гюйгенса еще недостаточно для решения дифракционных задач, если даже пользоваться точной формулировкой его, данной Кирхгофом (см. § 43). Этот принцип сводит только любую дифракционную задачу к определению волнового поля на произвольной замкнутой поверхности F, окружающей все источники света (см. рис. 150). Но точное определение поля на поверхности F возможно лишь после нахождения его во всем пространстве. Эту трудность Френель преодолел введением специальной гипотезы, которой мы, в сущности, уже пользовались в предыдущем параграфе. Сформулируем ее на примере непрозрачного плоского экрана с отверстиями, поставленного на пути распространяющейся волны.

Выберем в качестве вспомогательной поверхности F заднюю (т.е. неосвещаемую) сторону экрана. Примем, что на всех участках этой поверхности, которые прикрыты экраном, волновое поле равно нулю, а на отверстиях оно определяется приближенными законами геометрической оптики, т.е. поле такое, какое получилось бы в отсутствие экрана. Тем самым интегрирование в (39.1) распространяется только на отверстия, где волновое поле считается известным. Отметим недостатки введенной гипотезы.

1) Распространение гипотезы на неплоские экраны неоднозначно, так как в этом случае неясно, как провести вспомогательную поверхность через отверстия экрана.

2) Гипотеза противоречива. Если с помощью принципа Гюйгенса вычислить волновое поле во всем пространстве, то на поверхности F оно не совпадает с исходным полем, принятым при вычислении. На задних сторонах экранов вычисленное поле не обратится в нуль, а на отверстиях не будет совпадать с полем свободно распространяющейся волны.

3) Гипотеза допускает разрыв волнового поля на краях отверстий, что противоречит граничным условиям электродинамики Максвелла.

4) Гипотеза противоречит поперечности световых волн. Это противоречие не устраняется заменой скалярных колебаний векторными. Колебания, как и при изучении явлений интерференции, мы считали скалярными лишь ради простоты. При переходе к векторным колебаниям математическая формулировка принципа Гюйгенса и полученные из него результаты по существу не изменятся. Как строго показал Кирхгоф, для этого достаточно только скаляр E всюду заменить вектором $\mathbf{E}$.

Допустим теперь, что на непрозрачный экран с отверстием нормально падает плоская линейно поляризованная электромагнитная волна. На вспомогательной поверхности F вектор $\mathbf{E}$ будет иметь одно и то же направление, параллельное плоскости экрана. Принцип Гюйгенса сводит задачу о дифракции к суперпозиции коллинеарных векторных колебаний того же направления. Поэтому следует ожидать, что в дифрагированной волне вектор $\mathbf{E}$ всюду будет параллелен плоскости экрана. Это будет так и вдали от экрана, где дифрагированные волны разных направлений расходятся и перестают накладываться друг на друга. Так будет и в волне, дифрагировавшей косо к плоскости экрана. Но в действительности вектор $\mathbf{E}$ перпендикулярен к дифрагирующим лучам и образует с вычисленным направлением угол, равный «углу дифракции» ϑ (рис. 163).

Метод Френеля решения дифракционных задач может быть оправдан, когда размеры отверстий и непрозрачных промежутков между ними велики по сравнению с длиной световой волны, а потому заметная интенсивность света наблюдается лишь при малых углах дифракции. Действительно, в отверстиях законы геометрической оптики нарушаются лишь в непосредственной близи от их краев. Применяя гипотезу Френеля, мы при вычислении интеграла (39.1) пользуемся неправильными значениями подынтегральной функции только внутри узких полосок вблизи краев отверстий. Ширина этих полосок порядка длины световой волны. В основной области интегрирования используются правильные значения подынтегрального выражения. Автоматически снимается и возражение (4), связанное с поперечностью световых колебаний, так как различием направлений вычисленного и действительного полей при малых углах дифракции можно пренебречь. Все это в основном подтверждается совпадением результатов теоретических расчетов с опытом. Но для малых препятствий, сравнимых с длиной волны, и, следовательно, для больших углов дифракции гипотезой Френеля уже нельзя пользоваться.

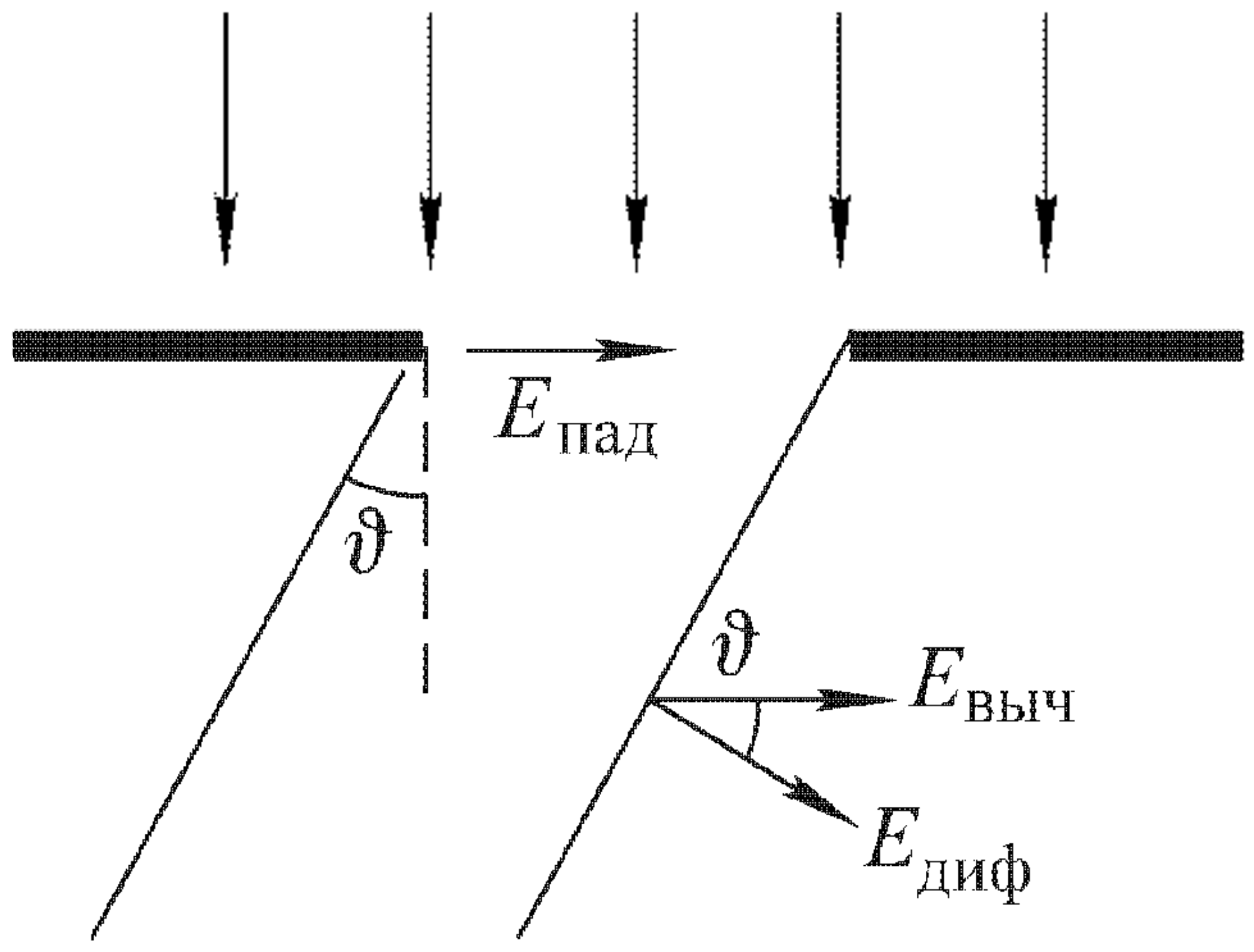


Рис. 163

Отметим еще, что гипотеза Френеля приводит к выводу о независимости дифрагированной волны от материала экрана. Этот вывод в основном также подтверждается опытом. Только более точные опыты обнаруживают и влияние материала экрана на дифракцию света.

2. Явления дифракции принято классифицировать в зависимости от расстояний источника и точки наблюдения (экрана) от препятствия, поставленного на пути распространения света. Если эти расстояния очень велики (бесконечно велики), то дифракция называется *дифракцией в параллельных лучах* или *дифракцией Фраунгофера* (1787–1826). В противоположном случае говорят о *дифракции в непараллельных лучах* или *дифракции Френеля.* Практически для осуществления дифракции Фраунгофера точечный источник света помещают в фокусе собирательной линзы. Получающийся параллельный пучок света дифрагирует на каком-то препятствии. Дифракционная картина наблюдается в фокальной плоскости линзы, поставленной на пути дифрагированного света, или в зрительную трубу, установленную на бесконечность. Ясно, что между фраунгоферовой и френелевой дифракциями нет принципиального различия и резкой границы. Одна непрерывно переходит в другую.

Для лучшего уяснения приведенной классификации начнем с примера. Рассмотрим круглое отверстие и точечный источник на его оси. Пусть сначала точка наблюдения также находится на оси. Если в отверстии укладывается небольшая часть первой зоны Френеля, то дифракция будет фраунгоферовой. В этом случае все колебания в плоскости отверстия совершаются и приходят в точку наблюдения практически в одинаковых фазах. При смещении точки наблюдения вбок появляются разности фаз между вторичными волнами, приходящими в точку наблюдения от различных точек отверстия. Этим и обусловлено появление дифракционных колец. Если отверстие заменить непрозрачным экраном, то этот случай, по соображениям, которые выяснятся в п. 4, также относят к дифракции Фраунгофера. Если же в отверстии или экране (для точки наблюдения, лежащей на оси системы) укладывается заметная часть первой зоны или несколько зон Френеля, то дифракция считается френелевой.

3. Рассмотрим теперь общий случай. Пусть препятствием является непрозрачный экран с отверстием (рис. 164), S — точечный источник монохроматического света, P — точка наблюдения. За начало координат примем произвольную точку O в плоскости отверстия. По принципу Гюйгенса–Френеля волновое поле в точке P представляется интегралом

$$E_P = \int \frac{1}{rr'} e^{i\Phi(\mathbf{R})}\, dF, \qquad (41.1)$$

где $\Phi = \omega t - k(r + r')$. Ввиду малости углов дифракции в подынтегральном выражении (41.1) опущен множитель $K(\alpha)$, определяющий зависимость поля вторичной волны от направления ее излучения. Знаменатель rr' можно считать постоянным, так как размеры отверстия предполагаются очень малыми как по сравнению с r, так и по сравнению с r'. Но зависимость фазы Φ от радиуса-вектора $\mathbf{R}$, определяющего положение элемента площади dF, конечно, должна быть учтена, так как фаза быстро меняется уже на малых расстояниях (порядка ширины френелевой зоны). Интегрирование в (41.1) во всех случаях выполняется по практически небольшому числу (не бо-

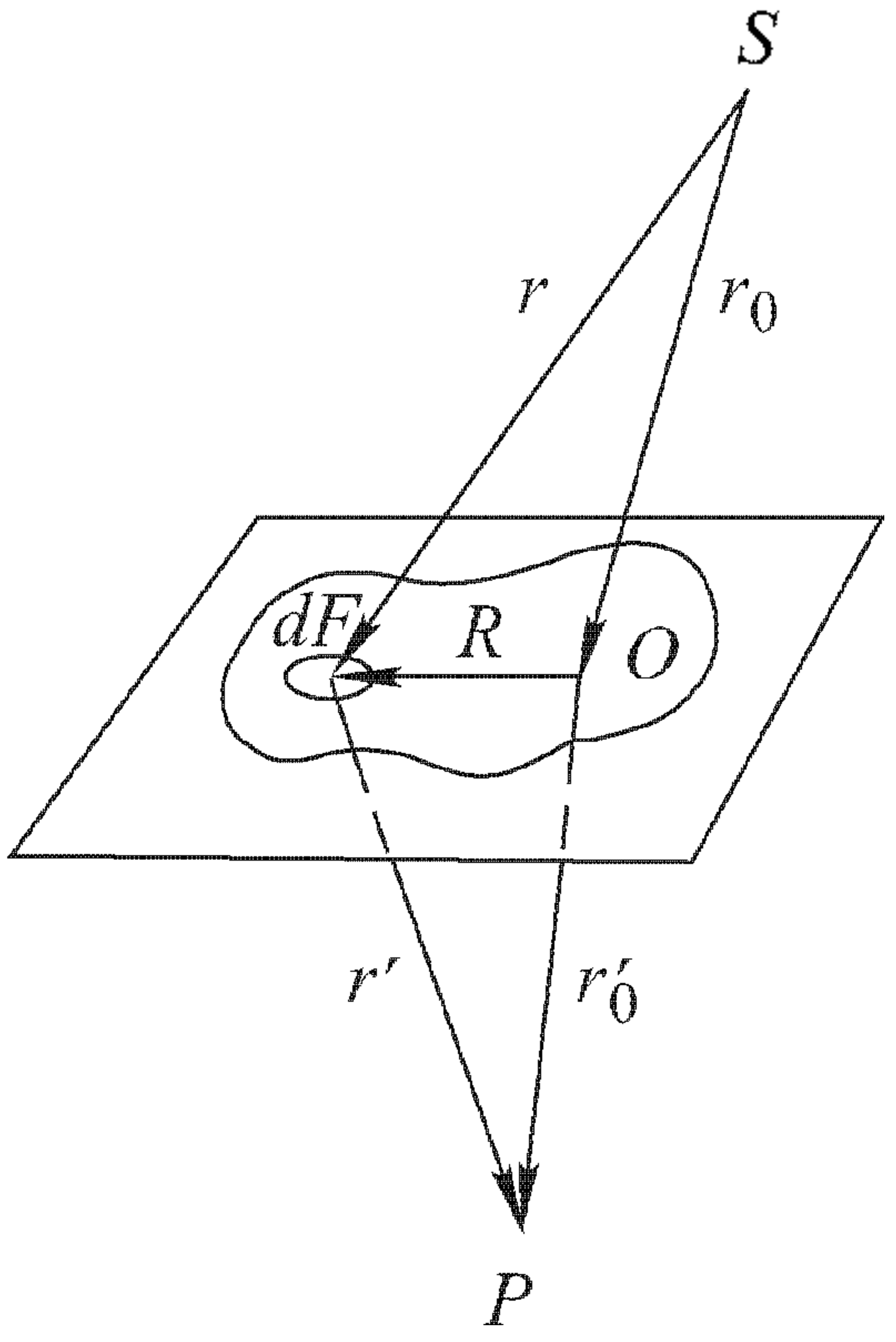


Рис. 164

лее нескольких десятков) полных или неполных зон Френеля (см., например, следующий параграф). Если расстояния r_0 и r_0' настолько велики, что разложение фазы $\Phi(\mathbf{R})$ по степеням $\mathbf{R}$ можно оборвать на членах первой степени, то такие расстояния можно считать бесконечно большими. В этом случае имеет место дифракция Фраунгофера. Если же такая точность недостаточна, то дифракция будет френелевой. Обычно в случае френелевой дифракции разложение фазы $\Phi(\mathbf{R})$ по степеням $\mathbf{R}$ достаточно оборвать на членах второй степени.

Найдем количественный критерий, при выполнении которого в фазе Φ можно пренебречь квадратичными членами и, следовательно, считать дифракцию фраунгоферовой. Как видно из рис. 164, $\mathbf{r} = \mathbf{r}_0 + \mathbf{R}$, откуда

$$r^2 = r_0^2 + 2(\mathbf{r}_0\mathbf{R}) + R^2.$$

Извлекая квадратный корень с точностью до квадратичных членов включительно, получим

$$r = r_0 + (\mathbf{sR}) + \frac{1}{2r_0}[R^2 + (\mathbf{sR})^2],$$

где через $\mathbf{s}$ обозначен единичный вектор $\mathbf{s} = \mathbf{r}_0/r_0$. Аналогично вычисляется расстояние r'. В результате получаем

$$\Phi = \omega t - k(r_0 + r_0') + k(\mathbf{s} - \mathbf{s}')\mathbf{R} + \\ + \frac{k}{2}\left[\left(\frac{1}{r_0} + \frac{1}{r_0'}\right)R^2 + \frac{1}{r_0}(\mathbf{sR})^2 + \frac{1}{r_0'}(\mathbf{s}'\mathbf{R})^2\right]. \quad (41.2)$$

В отличие от постоянного вектора $\mathbf{s}$, единичный вектор $\mathbf{s}' = \mathbf{r}_0'/r_0'$ меняется при изменении положения точки наблюдения. Этим и обусловлено появление дифракционных картин, т.е. изменение интенсивности светового поля от точки к точке. Слагаемое $-k(r_0 + r_0')$ не играет роли. Оно вносит в выражение для E_P только постоянный фазовый множитель $e^{-ik(r_0+r_0')}$, не влияющий на интенсивность светового поля в точке P. Квадратичные члены имеют порядки

$$kR^2/r_0 = 2\pi R^2/(\lambda r_0) \quad \text{и} \quad kR^2/r_0' = 2\pi R^2/(\lambda r_0').$$

Максимальные значения эти величины принимают при $R = D$, где D — диаметр отверстия. Практически этими членами можно пренебречь, если они меньше $\pi/2$. Действительно, характерной величиной для фазы является $\pi/2$, так как при изменении углов на такую величину синусы и косинусы претерпевают существенные изменения. Все слагаемые в фазе, много меньшие $\pi/2$, не играют роли и могут быть отброшены без особого ущерба для точности. Для максимальных изменений, обусловленных квадратичными членами, мы (несколько условно) заменим требование «много меньше» более слабым требованием «меньше».

Таким путем получим

$$r_0 > \frac{4D^2}{\lambda}, \quad r_0' > \frac{4D^2}{\lambda}. \tag{41.3}$$

Этими условиями и определяется область фраунгоферовой дифракции. Например, если $D = 1$ мм, $\lambda = 500$ нм, то фраунгоферова дифракция наступает при r_0, $r_0' > 4D^2/\lambda = 8$ м.

Легко также написать условия, при которых разложение фазы $\Phi(\mathbf{R})$ можно оборвать на членах второй степени:

$$r_0 > \sqrt{4D^3/\lambda}, \quad r_0' > \sqrt{4D^3/\lambda}. \tag{41.4}$$

В приведенном примере они удовлетворяются при r_0, $r_0' > 9$ см. Впрочем, квадратичным приближением (41.2) пользуются даже тогда, когда условия (41.4) не выполняются. Это делается при вычислении интеграла (41.1) в несущественных частях области интегрирования, где точное знание фазы Φ не имеет значения. Пример такого рода будет приведен в следующем параграфе при рассмотрении дифракции на прямолинейном крае экрана.

Теперь становится качественно ясным, как ведет себя световой пучок, прошедший через диафрагму, на различных расстояниях r от нее. Для определенности будем считать, что диафрагма освещается параллельным пучком света. На малых расстояниях $r \ll D^2/\lambda$ приближенно применима геометрическая оптика (прямолинейное распространение света). Затем с увеличением расстояния r наблюдается сложная картина френелевой дифракции. На больших расстояниях $r \gg D^2/\lambda$ дифракционная картина упрощается и переходит в дифракцию Фраунгофера.

4. Следствием гипотезы Френеля является полезная *теорема Бабине* (1794–1872) о свойствах так называемых *дополнительных экранов*. Возьмем плоский экран, освещаемый каким-либо источником света. Пусть $E_{\text{вх}}$ — поле падающей волны, какое получилось бы в точке (x, y) на передней поверхности экрана при отсутствии последнего, а $E_{\text{вых}}$ — поле в той же точке на задней поверхности экрана (экран предполагается тонким, и его толщина не играет роли). Пусть эти поля связаны соотношением $E_{\text{вых}} = \alpha_1 E_{\text{вх}}$. Коэффициент α_1 называется *пропускаемостью экрана.* Он может зависеть от координат x, y, структуры экрана и длины волны, но не зависит от напряженности волнового поля. Для другого экрана той же геометрической формы $E_{\text{вых}} = \alpha_2 E_{\text{вх}}$. Эти два экрана называются *дополнительными*, если $\alpha_1 + \alpha_2 = 1$. Примером могут служить два непрозрачных экрана с отверстиями, причем отверстия одного экрана совпадают с непрозрачными частями другого. Согласно (41.1), волновые поля в точке наблюдения при использовании дополнительных экранов могут быть записаны в виде

$$E_P^{(1)} = \int \frac{\alpha_1}{rr'} E_{\text{вх}} e^{i\Phi}\, dF, \qquad E_P^{(2)} = \int \frac{\alpha_2}{rr'} E_{\text{вх}} e^{i\Phi}\, dF,$$

причем в обоих случаях интегрирование распространяется по всей поверхности экрана. (Для определенности источник света предполагается точечным, что не имеет никакого значения.) Сложив эти два выражения и учтя, что $\alpha_1 + \alpha_2 = 1$, получим

$$E_P^{(1)} + E_P^{(2)} = \int \frac{1}{rr'} E_{\text{вх}} e^{i\Phi}\, dF.$$

Но последний интеграл есть поле E_P в точке P, которое получилось бы при свободном распространении волны. Таким образом,

$$E_P^{(1)} + E_P^{(2)} = E_P. \qquad (41.5)$$

Эта формула и выражает теорему Бабине в ее наиболее общей форме.

Обычно теорема Бабине понимается в более узком смысле — применительно к случаю фраунгоферовой дифракции. Пусть фраунгоферова дифракционная картина наблюдается в фокальной плоскости линзы. Если бы на пути параллельных лучей не было препятствий, то световое поле в этой плоскости было бы всюду равно нулю, за исключением фокуса линзы. Таким образом, согласно (41.5), во всякой точке фокальной плоскости, за исключением фокуса, должно быть $E_P^{(1)} + E_P^{(2)} = 0$. Так как интенсивность света I пропорциональна $|E_P|^2$, то отсюда получаем

$$I_1 = I_2. \qquad (41.6)$$

Следовательно, поскольку наблюдению доступна только интенсивность светового поля, а не его фаза, *фраунгоферовы дифракционные картины от дополнительных экранов, получаемые в фокальной плоскости линзы, всюду одинаковы, за исключением самого фокуса.*

ЗАДАЧА

На черный экран падает плоская световая волна. Из-за дифракции за экраном, наряду с неотклоненной волной, появятся волны всевозможных направлений (рассеянный свет). Показать, что количество рассеянной энергии равно количеству энергии, поглощенному экраном.

Решение. Заменим экран дополнительным, т.е. отверстием той же величины и формы. От этого по теореме Бабине интенсивность светового поля в бесконечности сохранится неизменной во всех направлениях, за исключением направления первичной волны. Но на любое строго фиксированное направление за отверстием приходится нулевая интенсивность света, так как отверстие рассеивает весь падающий на него свет. С другой стороны, экран по предположению полностью поглощает весь падающий свет. Отсюда непосредственно получается требуемый результат. Разумеется, он применим только в тех случаях, когда размеры экрана очень велики по сравнению с длиной волны.

§ 42. Зоны Шустера и спираль Корню

1. В одномерных задачах, например при рассмотрении дифракции на прямоугольной щели, разбиение волнового фронта на кольцевые зоны нецелесообразно. Лучше разбивать волновой фронт на *полосатые зоны*, называемые *зонами Шустера* (1851–1934). Ограничимся случаем, когда волновой фронт плоский, хотя обобщение на случай сферического фронта и не встречает никаких затруднений. Пусть плоскость волнового фронта AB перпендикулярна к плоскости рис. 165. Обозначим через b длину перпендикуляра PO, опущенного из точки наблюдения на волновой фронт. Проведем цилиндрические коаксиальные поверхности, ось которых проходит через точку P перпендикулярно к плоскости рисунка, а радиусы равны b, $b+\lambda/2$, $b+2(\lambda/2)$, … Тогда волновой фронт разобьется на прямоугольные полосы, которые и называются зонами Шустера. Центральную зону условимся считать за две зоны: одна расположена справа, а другая слева от точки O. Тогда $r_n^2 = b^2 + x_n^2$, $r_{n-1}^2 = b^2 + x_{n-1}^2$, а потому $r_n^2 - r_{n-1}^2 = x_n^2 - x_{n-1}^2$. Приближенно

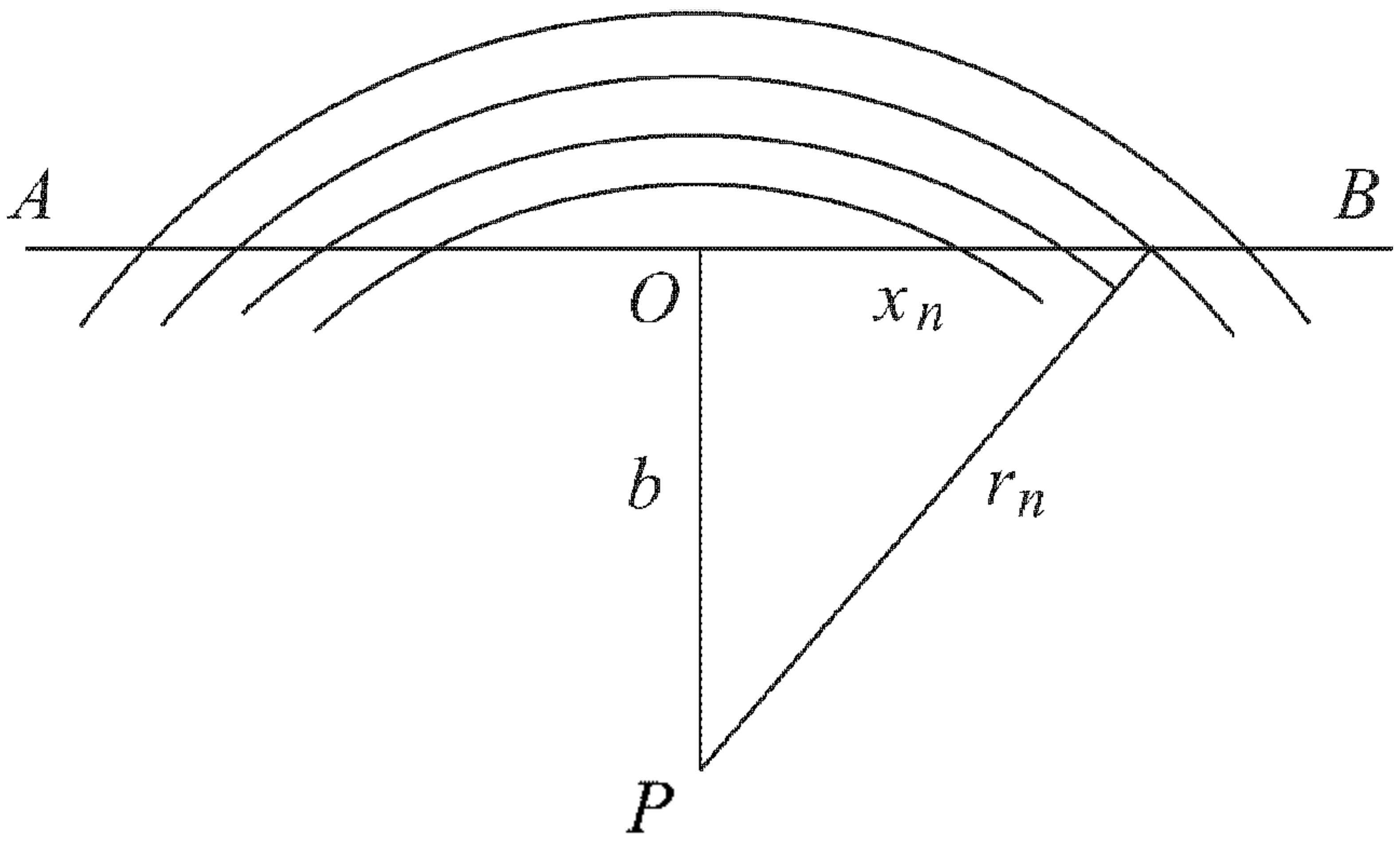


Рис. 165

$$r_n^2 - r_{n-1}^2 = (r_n + r_{n-1})(r_n - r_{n-1}) = 2b(\lambda/2) = b\lambda.$$

Таким образом, получаем рекуррентное соотношение

$$x_n^2 - x_{n-1}^2 = b\lambda, \tag{42.1}$$

из которого могут быть найдены все x_n. Так как $x_0 = 0$, то

$$x_1 = \sqrt{b\lambda}, \quad x_2 = \sqrt{2b\lambda}, \quad \ldots, \quad x_n = \sqrt{nb\lambda}. \tag{42.2}$$

Ширины последовательных зон Шустера будут

$$\sqrt{b\lambda}, \quad (\sqrt{2}-1)\sqrt{b\lambda}, \quad (\sqrt{3}-\sqrt{2})\sqrt{b\lambda}, \ldots \tag{42.3}$$

Они монотонно убывают и в пределе, когда $r \to \infty$, стремятся к $\lambda/2$, как это ясно из их построения. (Впрочем, высшие зоны не играют роли. Имеют значение только несколько десятков первых зон Шустера.)

Как и в случае зон Френеля, применим теперь графический метод (см. рис. 153). Каждую зону Шустера разобьем на узкие полоски и будем изображать колебание в точке P, вносимое отдельной полоской, вектором на векторной диаграмме. Затем перейдем к пределу, устремляя к нулю ширину каждой полоски. В результате получится плавная кривая, называемая *спиралью Корню* (1841–1902) (рис. 166). Она состоит из двух симметричных ветвей, бесконечное число раз обвивающихся вокруг «фокусов» F и F' и неограниченно приближающихся к ним. Верхняя ветвь представляет действие правой половины волнового фронта, нижняя — левой. Отличие каждой из ветвей от соответствующей спирали на рис. 153 обусловлено более быстрым убыванием начальных зон Шустера, чем зон Френеля. Колебание, возбуждаемое первой правой зоной Шустера, изображается вектором $\overrightarrow{OA}$, второй правой — вектором $\overrightarrow{A2}$, двумя первыми правыми зонами вместе — вектором $\overrightarrow{O2}$ и т.д. (все эти векторы на рис. 166 не проведены). Колебание, возбуждаемое всем волновым фронтом, представляется вектором $\overrightarrow{F'F}$, соединяющим фокусы спирали Корню. По мере приближения к фокусам амплитуды колебаний становятся все меньше и меньше и в пределе обращаются в нуль.

X

2. При нахождении уравнения спирали Корню надо учесть, что реально всегда приходится иметь дело не с бесконечными, а с ограниченными волновыми фронтами, причем заметная интенсивность наблюдается лишь при малых углах дифракции. Поэтому в формуле (41.1) изменения знаменателей r и r' (а также уже отброшенных ранее ослабляющих множителей $K(\alpha)$) можно не принимать во внимание. Если нас интересует только относительное распределение интенсивности, то можно положить $rr' = 1$. В плоскости волнового фронта фазу можно представить в виде $\Phi = \omega t - kr$ (здесь произведено переобозначение: в формуле (41.1) расстояние r обозначалось через r').

Примем волновой фронт за координатную плоскость XY, а начало координат поместим в точке O. Тогда $r^2 = b^2 + (x^2 + y^2)$, а следовательно, $r - b = (x^2 + y^2)/(2b) + \ldots$ Члены высших степеней можно отбросить, если даже они добавляют в фазу слагаемые порядка π и больше. Дело в том, что такие члены, как это видно из формы спирали Корню, не меняя общего характера дифракционной картины, производят в ней только практически незаметные смещения высших дифракционных максимумов и минимумов. Кроме того, высшие дифракционные максимумы и минимумы следуют друг за другом столь часто, что для их реального осуществления требуются точечные источники света высокой степени монохроматичности. В противном случае все дифракционные полосы высших порядков размываются и переходят в равномерно освещенный фон. Отбросим все фазовые множители, не влияющие на относительное распределение интенсивности светового поля. Тогда поле в точке наблюдения P представится интегралом

$$E_P = \iint e^{-ik(x^2+y^2)/(2b)}\, dx\, dy.$$

Интегрирование должно быть выполнено по всей открытой поверхности волнового фронта. Допустим, что в направлении оси Y она простирается достаточно далеко в обе стороны. Тогда интегрирование по y можно выполнить в пределах от $-\infty$ до $+\infty$, в результате чего появится постоянный множитель, не представляющий интереса. Интегрирование по x произведем от нуля, считая верхний предел x переменным (он может быть и положительным, и отрицательным). Вместо x, как это принято, введем новую переменную s по формуле $kx^2/b = \pi s^2$. Тогда

$$E_P = \int\limits_0^s e^{-i\pi s^2/2}\, ds, \tag{42.4}$$

$$E_P^* = \int\limits_0^s e^{i\pi s^2/2}\, ds. \tag{42.5}$$

При изображении колебаний можно пользоваться как выражением (42.4), так и комплексно сопряженным с ним (42.5). При построении спирали Корню обычно применяют выражение (42.5). Оно и представляет уравнение спирали Корню в комплексной форме. Если координатные оси выбраны так, как указано на рис. 166, то в прямоугольных координатах уравнение спирали Корню запишется в виде

$$X(s) = \int\limits_0^s \cos\frac{\pi s^2}{2}\,ds, \quad Y(s) = \int\limits_0^s \sin\frac{\pi s^2}{2}\,ds. \tag{42.6}$$

Входящие сюда интегралы называются *интегралами Френеля*. Очевидно,

$$X(s) = -X(-s), \quad Y(s) = -Y(-s),$$

т.е. кривая (42.6) симметрична относительно начала координат.

Полагая $s = \infty$, находим координаты фокусов спирали Корню:

$$X_F = Y_F = 1/2, \quad X_{F'} = Y_{F'} = -1/2.$$

Впрочем, для многих целей проще пользоваться непосредственно комплексной формой (42.5). В частности, для дифференциала дуги спирали Корню из (42.5) находим: $|e^{i\pi s^2/2}\,ds| = |ds|$. Отсюда следует, что параметр s есть длина дуги спирали, отсчитываемая от начала координат O.

Если τ — угол между касательной к спирали Корню и осью X, то $\operatorname{tg}\tau = dY/dX = \operatorname{tg}(\pi s^2/2)$, а потому

$$\tau = \pi s^2/2. \tag{42.7}$$

При $s = 0$ угол $\tau = 0$, т.е. в начале координат кривая касается оси X. При $s = 1$ касательная вертикальна и идет вверх. При $s = \sqrt{2}$, $\tau = \pi$ касательная снова горизонтальна, но идет в отрицательном направлении оси X. При $s = \sqrt{3}$, $\tau = {}^3\!/_2\pi$ она вертикальна и идет вниз. При $s = 2$, $\tau = 2\pi$ касательная принимает исходное — горизонтальное — направление. Формула (42.7) позволяет наглядно проследить, как кривая обвивается вокруг фокусов F и F', делая при этом бесконечное число оборотов. Эта формула особенно полезна в том отношении, что она позволяет по заданному параметру s легко находить соответствующую точку на спирали Корню.

Из формулы (42.7) получаем формулу для кривизны спирали Корню:

$$\frac{1}{R} = \frac{d\tau}{ds} = \pi s. \tag{42.8}$$

Длина всей спирали Корню бесконечна, а потому при приближении к фокусам ее кривизна стремится к бесконечности.

3. При работе со спиралью Корню надо знать значение параметра s. Его легко найти, зная на экране расстояние x точки наблюдения от центра картины O (см. рис. 165). Вычислив ширину первой зоны Шустера $\sqrt{\lambda b}$, находим далее $s = x\sqrt{2/(\lambda b)}$.

Рассмотрим в качестве примера дифракционную картину от прямолинейного края экрана (рис. 167). Где бы ни находилась точка наблюдения P, для нее всегда будет открыт правый край волнового фронта. На векторной диаграмме (см. рис. 166) колебание в точке наблюдения представится вектором $\overrightarrow{M_nF}$, конечная точка которого всегда находится в верхнем фокусе F, а начальная M_n лежит где-то на спирали Корню. Если, сохраняя неизменным положение конечной точки F, перемещать начальную точку M_n вдоль спирали Корню (положения M_1, M_2, M_3, ...), то таким путем можно получить распределение амплитуд и интенсивности колебаний света по всему экрану.

Обозначим через $a_0 = |FF'|$ и $I_0 = a_0^2$ амплитуду и интенсивность волны, когда открыт весь волновой фронт. Когда точка наблюдения P находится на границе геометрической тени, то колебание представится вектором $\overrightarrow{OF} = {}^1\!/_2\overrightarrow{F'F}$. Ему соответствует амплитуда $a_0/2$ и интенсивность $I_0/4$. При перемещении точки P в освещенную область экрана изображающая точка M_n начнет перемещаться по нижней ветви спирали Корню, а амплитуда и интенсивность колебаний будут последовательно проходить через максимумы и минимумы. Максимальная амплитуда, как видно из рис. 166, составляет $1{,}12a_0$, а интенсивность $1{,}25I_0$; минимальные значения их соответственно $0{,}89a_0$ и $0{,}78I_0$. При дальнейшем продвижении в освещенную область интенсивность асимптотически приближается к I_0. При погружении точки P в область геометрической тени изображающая точка M_n перемещается по верхней ветви спирали Корню. При этом по мере погружения в указанную область интенсивность света монотонно убывает и асимптотически стремится к нулю.

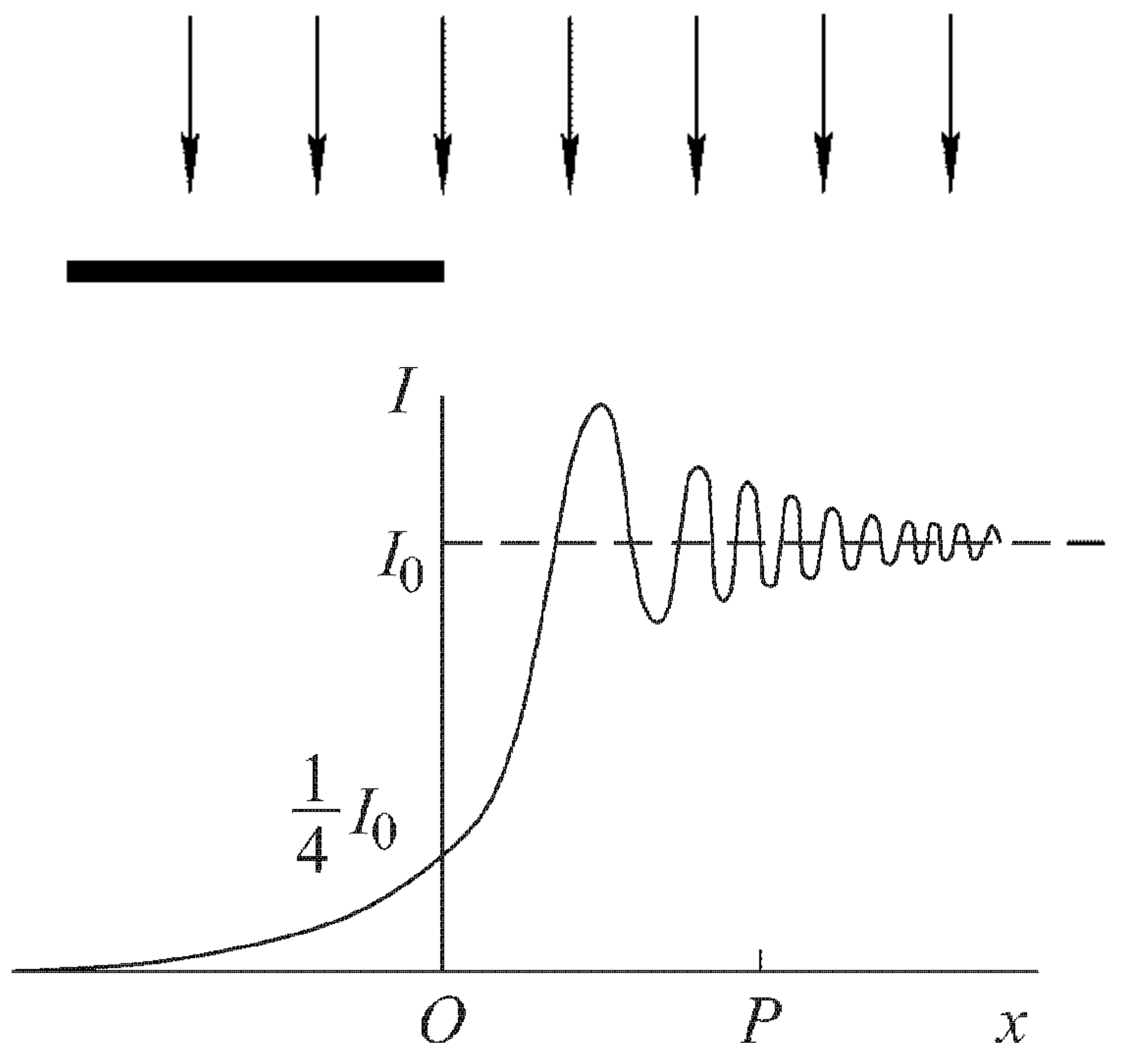


Рис. 167

Распределение интенсивности графически представлено на рис. 167. Таким образом, нет резкой границы между светом и тенью: в области геометрической тени интенсивность света убывает непрерывно и монотонно, а освещенная область расщепляется в дифракционные полосы. На рис. 168 показана дифракционная картина, наблюдаемая при дифракции света на крае экрана. Таким же путем можно рассчитать дифракционную картину на щели или длинном прямоугольном экране. На рис. 169 показана тень проволоки от точечного (или линейного) источника.

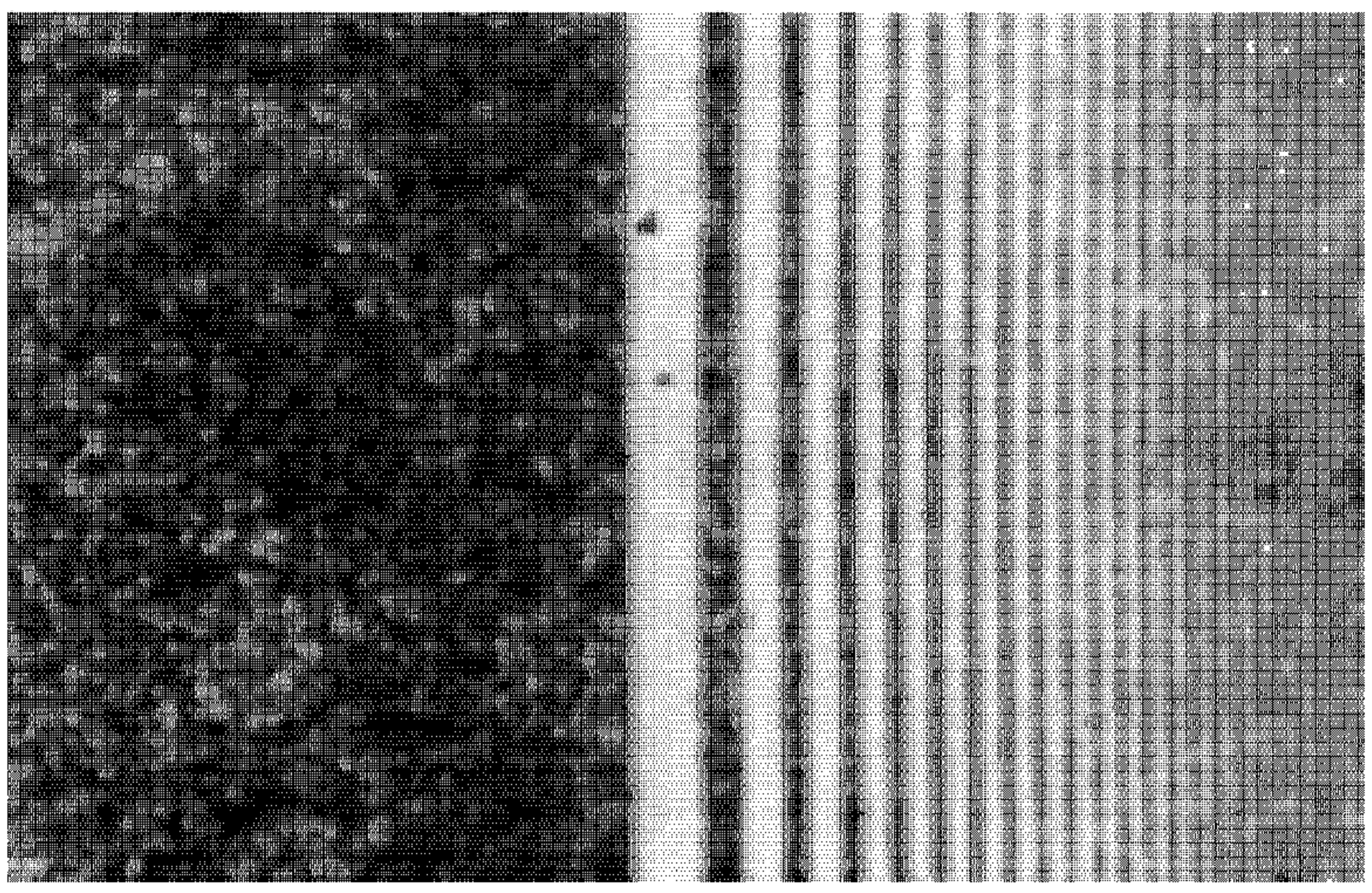


Рис. 168

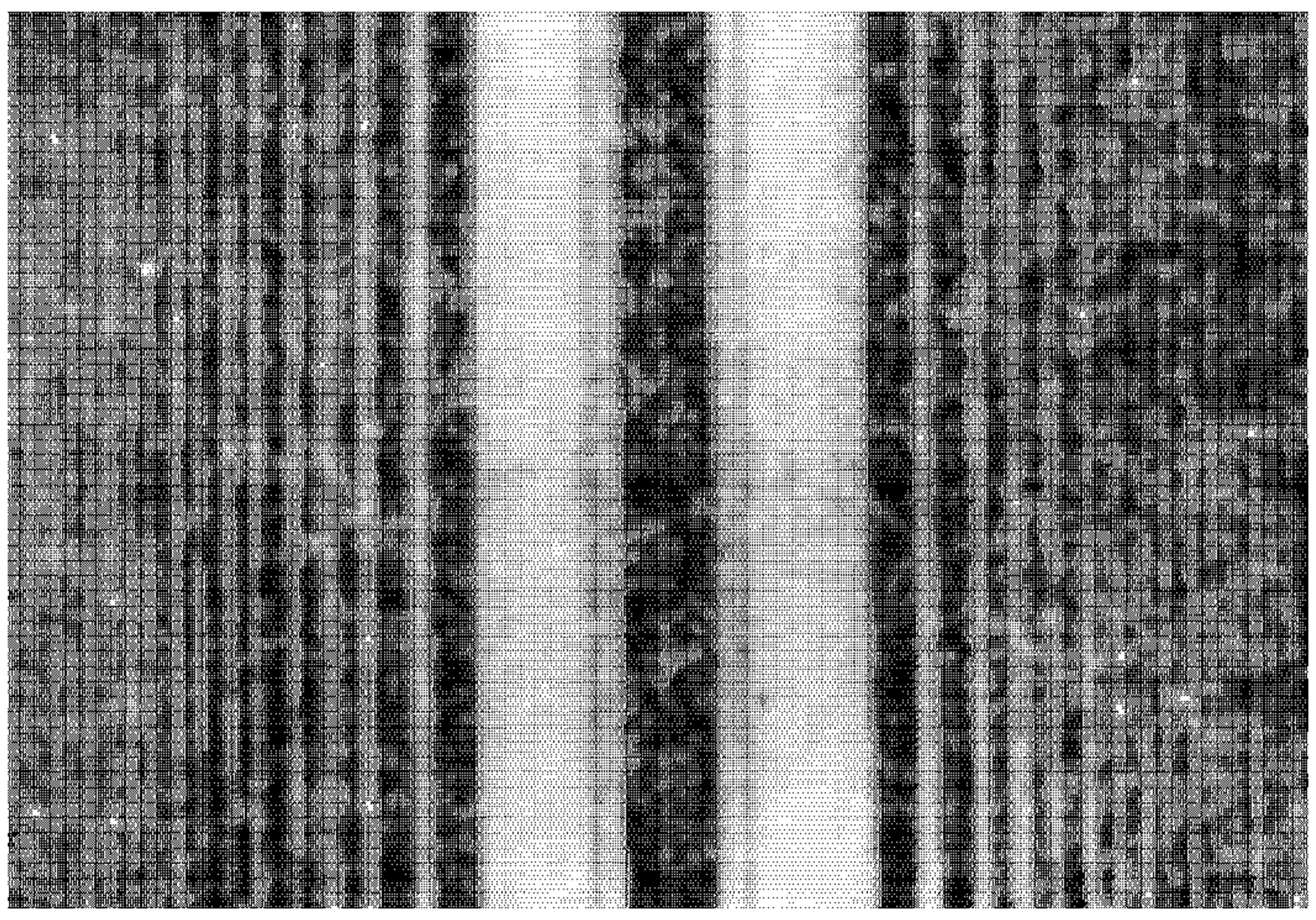


Рис. 169

ЗАДАЧИ

1. Положение светлых и темных полос при дифракции плоской волны на крае экрана можно приближенно (но с достаточной точностью) определить по точкам пересечения нижней ветви спирали Корню с прямой $F'F$, соединяющей ее фокусы, приняв во внимание, что в этих точках прямая FF' практически перпендикулярна к спирали Корню. Найти таким путем координаты указанных полос.

Ответ.

$$x_n = {}^1\!/_2\sqrt{\lambda r(8n-5)} \quad \text{(светлые полосы)},$$
$$x_n = {}^1\!/_2\sqrt{\lambda r(8n-1)} \quad \text{(темные полосы)},$$

где $n = 1, 2, 3, \ldots$

2. Когда звезда проходит мимо края Луны, получаются дифракционные полосы. Определить скорость их движения v по земной поверхности и оценить порядок их ширины Δx. Для наблюдения полос можно воспользоваться телескопом, в фокусе которого помещен фотоэлемент. Полосы, проходя перед объективом телескопа, возбуждают переменные электрические токи, которые могут быть усилены и с помощью осциллографа записаны на движущейся ленте. Оценить длительность t прохождения полос перед объективом.

Принимая звезду за равномерно светящийся диск, можно теоретически рассчитать распределение освещенности в дифракционной картине. Сравнивая его с распределением освещенности, найденным экспериментально, можно определить угловой диаметр звезды, что и было фактически выполнено в 1946 г. Уайтфордом для четырех звезд на стодюймовом рефлекторе Маунт-Вильсоновской обсерватории. Оценить угловые размеры звезд, для которых может быть пригоден этот метод.

Ответ. $v \approx 500$ м/с; $\Delta x \sim \sqrt{b\lambda} \approx 10$ м, где $b = 3{,}8 \cdot 10^5$ км — среднее расстояние до Луны. Метод пригоден, когда угловые размеры звезд лежат примерно в интервале от 10^{-4} до 10^{-3} угловой секунды.

3. Для получения фотографий дифракционных картин в тех случаях, когда источник света и экран расположены очень далеко, В.К. Аркадьевым (1884–1953) был применен *метод подобия*, в котором вместо действительных препятствий, стоящих на пути лучей, используются их уменьшенные и подобные модели, но длина волны сохраняется неизменной.

Требуется получить фотографию дифракционной картины от диска диаметром $D = 50$ см, когда на его оси расположен точечный источник света на расстоянии $A = 25$ км, а экран удален от него на $B = 50$ км (плоскость экрана перпендикулярна к оси диска). С этой целью диск заменили уменьшенной моделью с диаметром $d = 1$ см. Пользуясь методом зон Френеля, определить, на каких расстояниях a и b следует поместить источник света и экран, чтобы получилась подобная и уменьшенная в $n = 50$ раз дифракционная картина.

Ответ. $a = A/n^2 = 10$ м, $b = B/n^2 = 20$ м.

§ 43. Принцип Гюйгенса в формулировке Кирхгофа

1. До Кирхгофа принцип Гюйгенса–Френеля оставался гипотезой. Кирхгоф в 1883 г. вывел формулу, которую можно рассматривать как уточненную формулировку указанного принципа. Приведем вывод формулы Кирхгофа, хотя в дальнейшем и не будем ею пользоваться. Читатель может опустить его без ущерба для понимания дальнейшего.

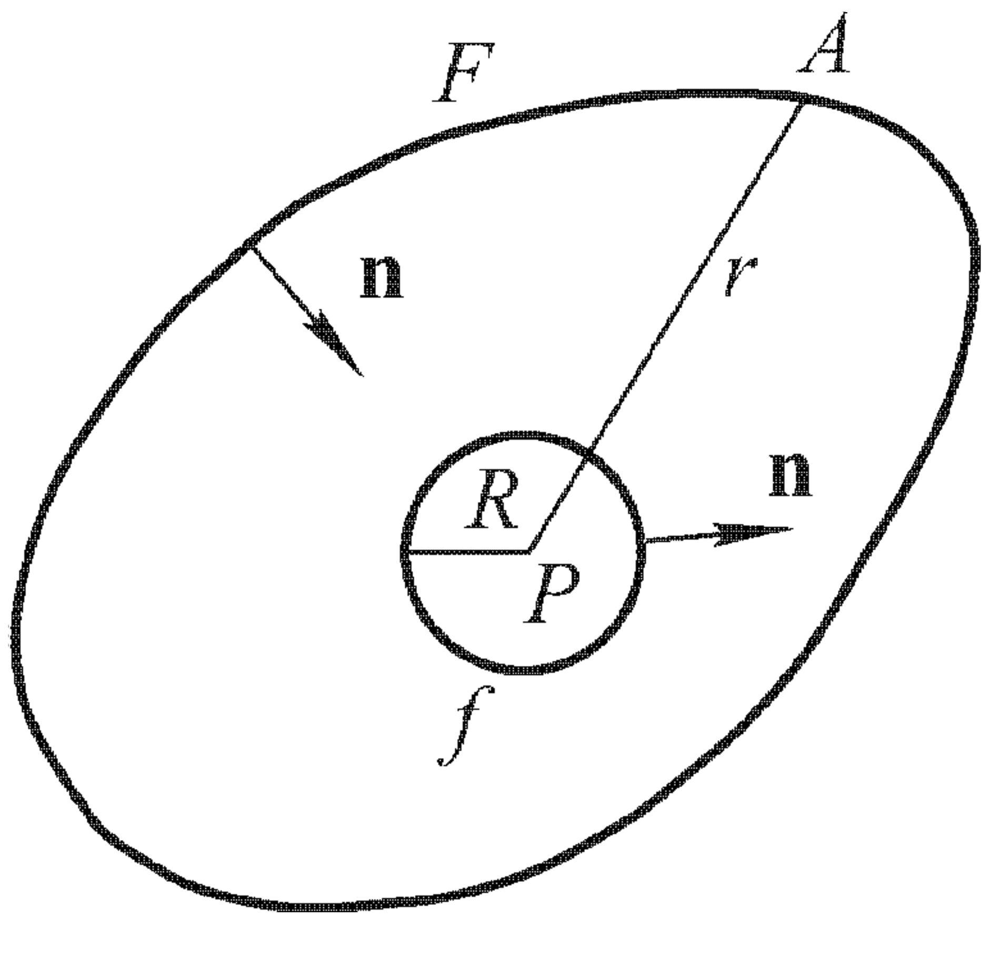


Рис. 170

Допустим, что среда, в которой распространяется свет, однородна. Будем характеризовать световое поле какой-то величиной E. Под E можно понимать либо вектор $\mathbf{E}$, либо вектор $\mathbf{B}$, либо одну из их проекций на декартовы оси координат. Эта величина удовлетворяет волновому уравнению

$$\Delta E - \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2} = 0,$$

которое в случае монохроматического поля переходит в

$$\Delta E + k^2 E = 0. \tag{43.1}$$

Найдем значение E в произвольной точке пространства P (рис. 170). Обозначим через r переменное расстояние какой-либо точки A от P. Величина

$$\chi = \frac{1}{r} e^{-ikr}, \tag{43.2}$$

рассматриваемая как функция точки A, также удовлетворяет уравнению

$$\Delta\chi + k^2\chi = 0. \tag{43.3}$$

Окружим точку P произвольной замкнутой поверхностью F, и притом такой, что в окружаемом ею пространстве нет источников света. Функция χ обращается в бесконечность в точке P. Исключим эту точку, окружив ее сферой f достаточно малого радиуса R с центром в P. Тогда во всем пространстве между сферой f и поверхностью F функции E и χ, а также их производные будут конечны и непрерывны. К ним можно применить формулу Грина (1793–1841)

$$\int (E\Delta\chi - \chi\Delta E)\,dV = -\int\limits_{F+f}\left(E\frac{\partial\chi}{\partial n} - \chi\frac{\partial E}{\partial n}\right)dF, \tag{43.4}$$

где V — объем пространства между поверхностями f и F, а $\mathbf{n}$ — внутренняя нормаль по отношению к этому пространству. Так как в указанном пространстве источников света нет, то в нем справедливы уравнения (43.1) и (43.3), Следовательно,

$$E\,\Delta\chi - \chi\,\Delta E = -k^2(E\chi - \chi E) = 0,$$

а потому

$$\oint\limits_f \left(E\frac{\partial\chi}{\partial n} - \chi\frac{\partial E}{\partial n}\right)df = -\oint\limits_F \left(E\frac{\partial\chi}{\partial n} - \chi\frac{\partial E}{\partial n}\right)dF.$$

Перейдем в этом равенстве к пределу, устремляя радиус сферы f к нулю. Правая часть равенства при этом не будет меняться. Что касается левой, то, взяв радиус R настолько малым, чтобы $kR \ll 1$, можно заменить экспоненциальный множитель e^{-ikR} единицей. Тогда левая часть примет вид

$$\oint\limits_f \left[E\frac{\partial}{\partial R}\left(\frac{1}{R}\right) - \frac{1}{R}\frac{\partial E}{\partial R}\right]df.$$

Так как величины E и $\dfrac{\partial E}{\partial R}$ в окрестности точки P конечны, то интеграл от второго слагаемого будет порядка $-4\pi R\,\dfrac{\partial E}{\partial R}$, т.е. при $R \to 0$ он обратится в нуль. Интеграл же от первого слагаемого в пределе перейдет в $-4\pi E_P$. Окончательно имеем

$$E_P = \frac{1}{4\pi}\oint\limits_F \left[E\frac{\partial}{\partial n}\left(\frac{e^{-ikr}}{r}\right) - \frac{e^{-ikr}}{r}\frac{\partial E}{\partial n}\right]dF, \tag{43.5}$$

где E_P — значение функции E в точке P.

2. Формула (43.5) по виду напоминает принцип Гюйгенса–Френеля. И тут и там поле в точке P выражается интегралом по замкнутой поверхности F. Однако у Френеля источники света лежат *внутри* замкнутой поверхности F, а точка P *вне* этой поверхности. Формула же (43.5), наоборот, предполагает, что точка P лежит *внутри* поверхности F, а источники *вне* ее. Легко, однако, преобразовать формулу (43.5), чтобы указанное различие исчезло. Для этого предположим, что все источники света S_1, S_2, S_3, … лежат в конечной области пространства. Окружим эту область замкнутой поверхностью F (рис. 171). Пусть точка P находится в пространстве вне поверхности F. Опишем из P как из центра сферу f настолько большого радиуса, чтобы она целиком окружала поверхность F. Тогда в пространстве между f и F не будет источников света, а потому можно для вычисления E в точке P применить формулу (43.5):

$$E_P = \frac{1}{4\pi} \int\limits_{F+f} \left(E\,\frac{\partial \chi}{\partial n} - \chi\,\frac{\partial E}{\partial n} \right) dF.$$

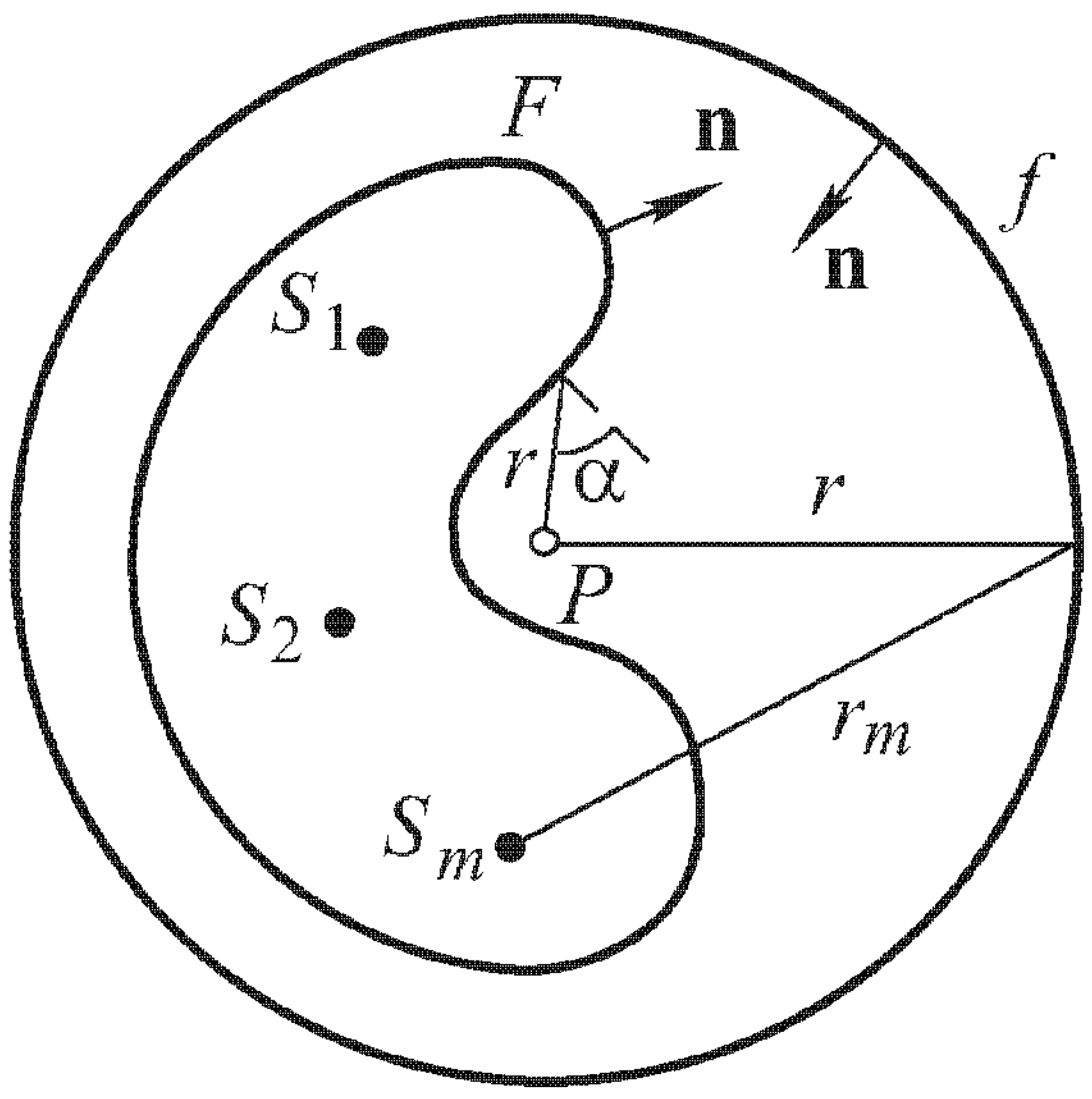


Рис. 171

Докажем, что интеграл по сфере f стремится к нулю, когда ее радиус стремится к бесконечности. Для этого необходимо выяснить поведение функции E на бесконечности. Предположим, что в пространстве, ограниченном F, находится один или несколько точечных источников света: S_1, S_2, S_3, … Тогда поле этих источников представится в виде

$$E = \sum C_m \frac{e^{-\imath k r_m}}{r_m},$$

где C_m — постоянные коэффициенты. Если r стремится к ∞, то r_m также стремится к ∞, однако разность $r_m - r$ будет оставаться конечной. Представим E в виде

$$E = \sum C_m \frac{e^{-\imath k(r+a_m)}}{r+a_m} = \frac{e^{-\imath kr}}{r} \sum \frac{A_m}{1 + a_m/r},$$

где A_m — новые постоянные. Разлагая выражение под знаком суммы в ряд по степеням $1/r$, получим

$$E = \frac{e^{-\imath kr}}{r} \left\{ \sum A_m - \frac{1}{r} \sum A_m a_m + \frac{1}{r^2}(\ldots) + \ldots \right\},$$

или

$$E = \frac{e^{-\imath kr}}{r}(C + \Phi) = (C + \Phi)\chi,$$

где C — постоянная, а Φ стремится к нулю по крайней мере как $1/r$. Подставляя это значение E в интеграл по сфере f, получим

$$\oint\limits_f \left(E\,\frac{\partial \chi}{\partial n} - \chi\,\frac{\partial E}{\partial n} \right) df = -\oint\limits_f \chi^2 \frac{\partial \Phi}{\partial n}\, df.$$

Подынтегральное выражение в последнем интеграле стремится к нулю по крайней мере как $1/r^3$, тогда как поверхность сферы обращается в бесконечность как r^2. Поэтому при $r \to \infty$ весь интеграл стремится к нулю. Таким образом, если сферу f удалить в бесконечность, то получится

$$E_P = \frac{1}{4\pi} \oint\limits_F \left[E\,\frac{\partial}{\partial n}\left(\frac{e^{-\imath kr}}{r}\right) - \frac{e^{-\imath kr}}{r}\,\frac{\partial E}{\partial n} \right] dF. \tag{43.6}$$

Эти рассуждения приводят также к следующему важному результату. Если некоторый участок поверхности F удаляется в бесконечность, то часть интеграла (43.6) по этому участку стремится к нулю. При этом предполагается, что все источники света находятся в *конечной области пространства.*

3. Формулы (43.5) и (43.6) и выражают *принцип Гюйгенса в формулировке Кирхгофа.* В обеих формулах $\mathbf{n}$ означает внутреннюю нормаль по отношению к тому пространству, в котором находится точка наблюдения P.

Выполнив дифференцирование по $\mathbf{n}$ и приняв во внимание, что $\partial r/\partial n = -\cos\alpha$, где α — угол между нормалью $\mathbf{n}$ и направлением из площадки dF на точку P, получим

$$E_P = \oint K(\alpha, r)\frac{e^{-ikr}}{r}\,dF. \tag{43.7}$$

Здесь введено обозначение

$$K(\alpha, r) = \frac{1}{4\pi}\left[\left(ik + \frac{1}{r}\right)E\cos\alpha - \frac{\partial E}{\partial n}\right]. \tag{43.8}$$

Тем самым установлена связь формулы Кирхгофа с принципом Гюйгенса: подынтегральное выражение в формуле (43.8) может рассматриваться как *вторичная волна*, распространяющаяся от площадки dF к точке P. Множитель K, однако, зависит *не только от угла* α, как предполагал Френель, но также и *от расстояния* r. В противном случае вторичная волна не могла бы удовлетворять волновому уравнению. Таким образом, вторичные волны не обладают шаровой симметрией. Они сферические только в том смысле, что их волновые фронты имеют форму сфер. Амплитуды же зависят от направления распространения и меняются с расстоянием иначе, чем $1/r$. Только в «волновой зоне», когда расстояние точки P от излучающего центра dF очень велико по сравнению с длиной волны, можно в выражении (43.8) пренебречь $1/r$ по сравнению с ik. Тогда

$$E_P = \frac{1}{4\pi}\oint\left(ikE\cos\alpha - \frac{\partial E}{\partial n}\right)\frac{e^{-ikr}}{r}\,dF. \tag{43.9}$$

Благодаря малости длин световых волн такой упрощенной формой принципа Гюйгенса в оптике можно пользоваться при решении всех конкретных задач.

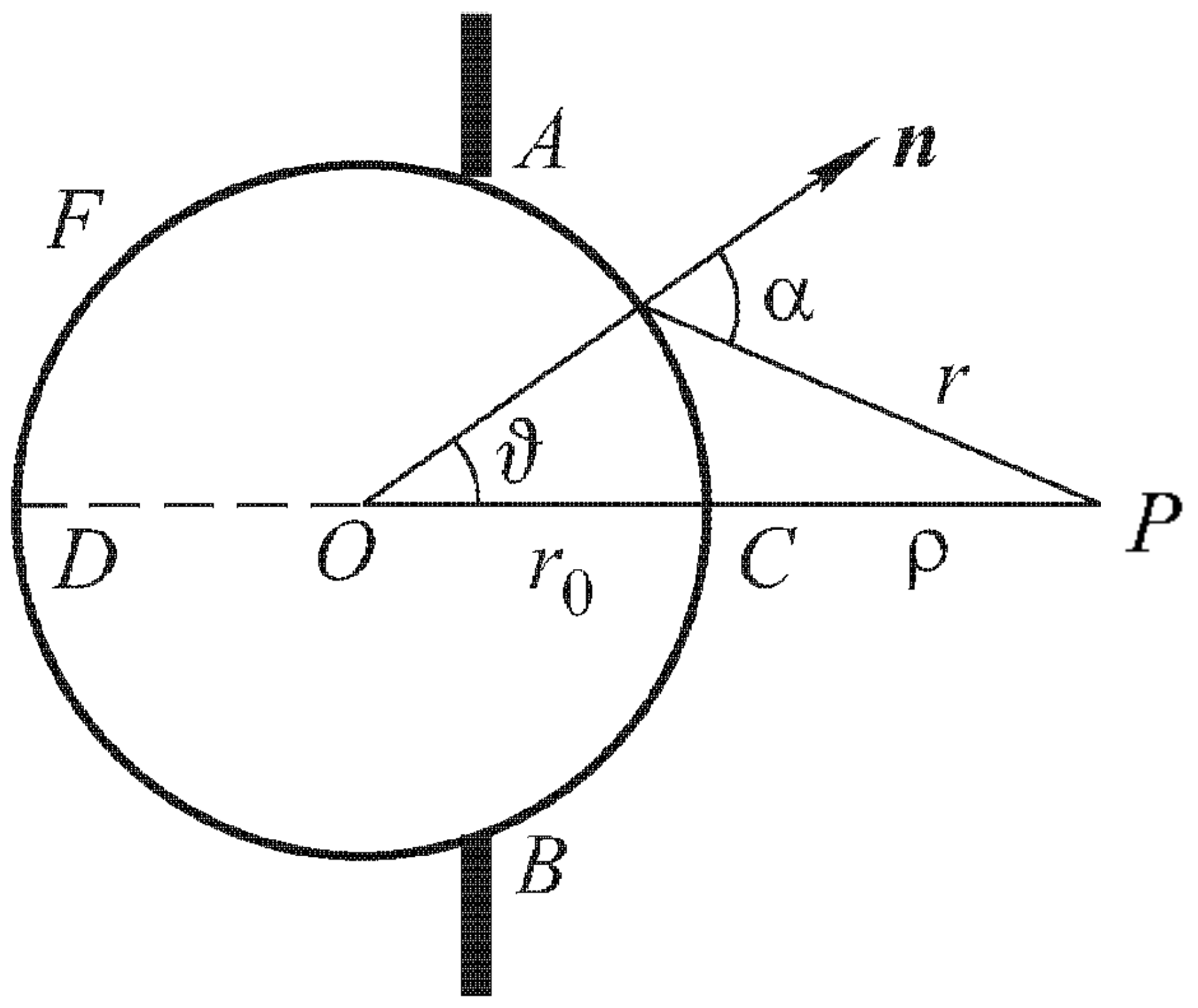


Рис. 172

4. Чтобы составить на примере более конкретное представление о вторичных волнах, рассмотрим свободное распространение сферической волны от точечного источника. В качестве поверхности F возьмем сферу радиуса r_0 с центром в источнике O (рис. 172). Поле на поверхности F представим выражением

$$E_0 = \frac{e^{i(\omega t - kr_0)}}{r_0}.$$

Предполагая, что радиус r_0 очень велик по сравнению с длиной волны, отсюда найдем $\partial E/\partial n = \partial E_0/\partial r_0 = -ikE_0$. Подставляя это значение в формулу (43.9), получим

$$E_P = \frac{ik}{4\pi}\oint(1 + \cos\alpha)E_0\frac{e^{-ikr}}{r}\,dF. \tag{43.10}$$

Сравнение этой формулы с формулой (43.7) дает

$$K(\alpha) = \frac{ik}{4\pi}(1+\cos\alpha). \tag{43.11}$$

Это и есть «ослабляющий множитель» $K(\alpha)$, введенный в § 39 ad hoc. Из теории автоматически получается, что он чисто мнимый и с возрастанием α монотонно убывает по абсолютной величине. Он обращается в нуль при $\alpha = \pi$, т.е. в точке D сферы, диаметрально противоположной точке наблюдения P.

Интеграл (43.10) теперь можно вычислить, поскольку он не содержит никаких неизвестных функций. Произведем это вычисление, так как таким путем можно получить строгое обоснование метода зон Френеля и результатов, полученных этим методом. Взяв за переменную интегрирования r и использовав значение E_0, преобразуем интеграл (43.10) к виду

$$E_P = \int\limits_{r_1}^{r_2} A(r)e^{-ikr}\,dr, \tag{43.12}$$

где введено обозначение

$$A(r) = \frac{ik(1+\cos\alpha)}{2(\rho+r_0)}e^{i(\omega t-kr_0)}, \tag{43.13}$$

а через r_1 и r_2 обозначены наибольшее и наименьшее значения, принимаемые r. Интегрируя (43.12) по частям, получим

$$E_P = -\frac{A(r)e^{-ikr}}{ik}\bigg|_{r_1}^{r_2} + \frac{1}{ik}\int\limits_{r_1}^{r_2}\frac{dA}{dr}e^{-ikr}\,dr. \tag{43.14}$$

Здесь единственной величиной, зависящей от r, является $\cos\alpha$. Найдем ее производную. Из рис. 172 следует, что $(\rho+r_0)^2 = r^2 + r_0^2 + 2rr_0\cos\alpha$, откуда $\dfrac{d\cos\alpha}{dr} = -\left(\dfrac{1}{r_0} + \dfrac{\cos\alpha}{r}\right)$. Взяв от (43.13) логарифмическую производную по r и воспользовавшись предыдущей формулой, найдем

$$\frac{1}{ik}\frac{dA}{dr} = \frac{i}{1+\cos\alpha}\left(\frac{1}{kr_0} + \frac{\cos\alpha}{kr}\right)A(r).$$

Такой величиной, а следовательно, и интегралом в формуле (43.14) следует пренебречь, так как это уже было сделано при выводе исходной формулы (43.9). Таким образом, в формуле (43.14) остается только первое слагаемое, т.е.

$$E_P = \frac{A(r_1)e^{-ikr_1} - A(r_2)e^{-ikr_2}}{ik}. \tag{43.15}$$

Существенно, что величина E_P представляется разностью одной и той же функции, но при различных значениях аргумента r, причем при изменении r на $\lambda/2$ знак этой функции меняется на противоположный. Этого достаточно для обоснования основного результата (39.6), на котором основан метод зон Френеля. Действительно, применив формулу (43.15) к первой зоне Френеля, найдем, что действие этой зоны может быть представлено в виде $E_1 = (u_1+u_2)$, действие второй зоны — в виде $E_2 = -(u_2+u_3)$, и т.д. Явный вид выражений u_i для доказательства не имеет значения. Действие первых N зон выразится суммой

$$E = (u_1 + u_2) - (u_2 + u_3) + \ldots + (-1)^{N+1}(u_N + u_{N+1}),$$

т.е.

$$E = u_1 + (-1)^{N+1}u_{N+1}.$$

Так как по величине действия двух соседних зон почти одинаковы, то $u_{N+1} = u_N$. С той же степенью точности $^1\!/_2 E_1 = u_1$, $^1\!/_2 E_N = (-1)^{N+1}u_N = (-1)^{N+1}u_{N+1}$. Следовательно,

$$E = \frac{1}{2}(E_1 + E_N).$$

§ 44. Дифракция Фраунгофера на щели

1. Дифракционные явления Фраунгофера имеют в оптике значительно большее практическое значение, чем дифракционные явления Френеля. При практическом осуществлении дифракции Фраунгофера источник света S помещается в фокусе линзы (рис. 173). (Линза L_1 не нужна, если источником света служит лазер, поскольку от него исходит уже параллельный пучок света.) Дифракция возникает на каком-либо препятствии AB, поставленном на пути световых лучей, прошедших через линзу L_1. Дифракционная картина наблюдается в фокальной плоскости другой линзы L_2. Ее можно также наблюдать в зрительную трубу, установленную на бесконечность. Но в теоретических рассуждениях удобнее от всех этих вспомогательных приспособлений отвлечься, предполагая, что на препятствие AB падает параллельный пучок лучей, а дифракция наблюдается «в бесконечности».

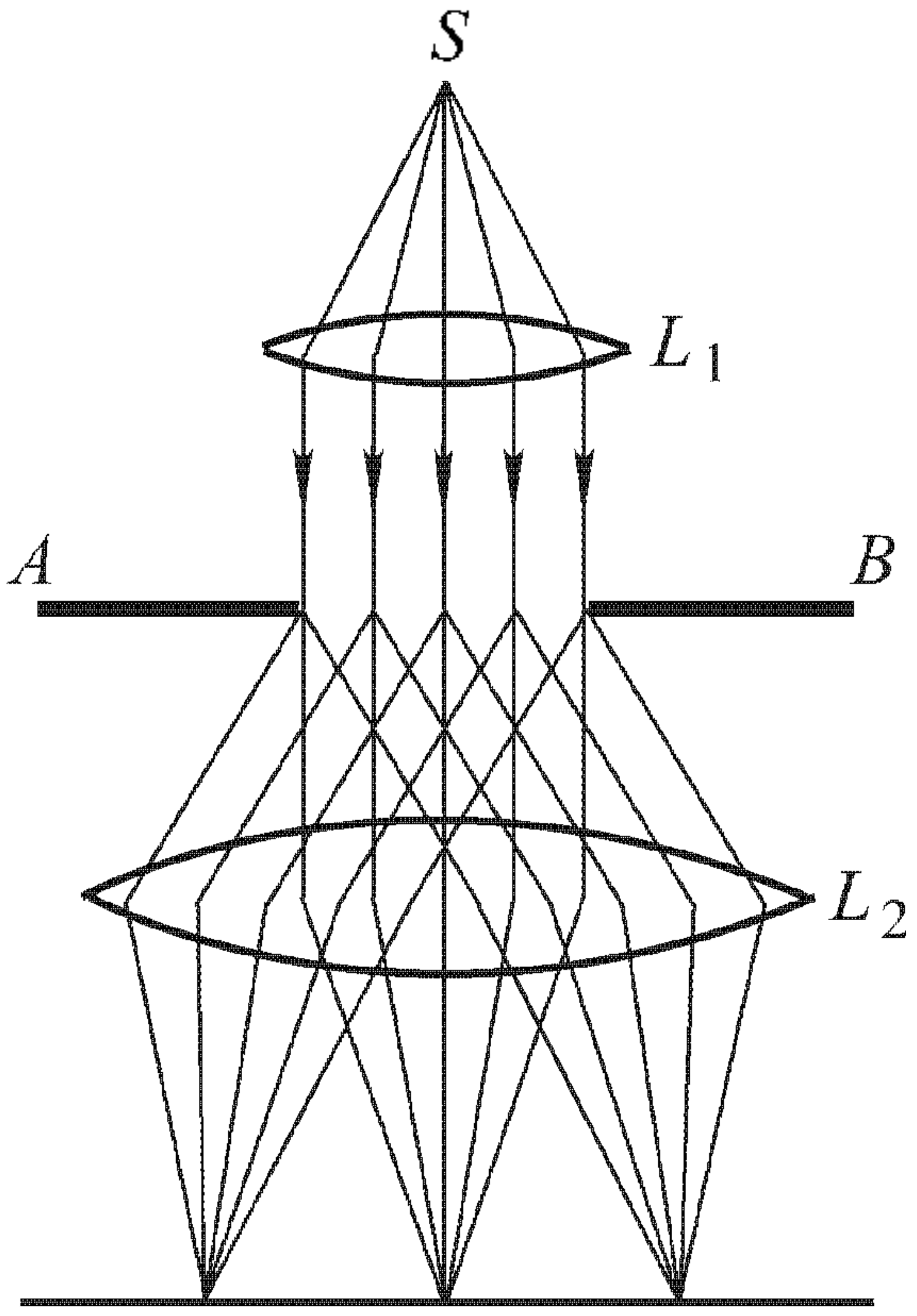


Рис. 173

Простейшим для расчета и практически очень важным случаем является фраунгоферова дифракция на длинной прямоугольной щели. Ширину щели обозначим через b, ее длину будем считать бесконечной. Пусть на щель нормально падает плоская монохроматическая волна (рис. 174). Световое поле за щелью найдется по принципу Гюйгенса как результат интерференции когерентных вторичных волн, исходящих из различных точек волнового фронта на щели. Вторичные волны, излучаемые полоской волнового фронта ширины dx, параллельной щели, складываясь, дают цилиндрическую волну, осью которой является эта полоска. Зависимость этой волны от направления ее распространения, определяемого углом ϑ, можно не учитывать, так как задача решается методом Френеля, а потому угол дифракции ϑ должен предполагаться малым. Однако необходимо учесть разности фаз между волнами, исходящими из различных полосок. Разумеется, речь идет о фазах колебаний на бесконечном расстоянии от щели. Волна, исходящая из dx под углом ϑ, опережает по фазе волну того же направления, исходящую из середины щели O, на $kx\sin\vartheta$. Поэтому результирующее поле в бесконечности, создаваемое всей щелью, представится интегралом

$$E = \int\limits_{-b/2}^{+b/2} e^{ikx\sin\vartheta}\,dx.$$

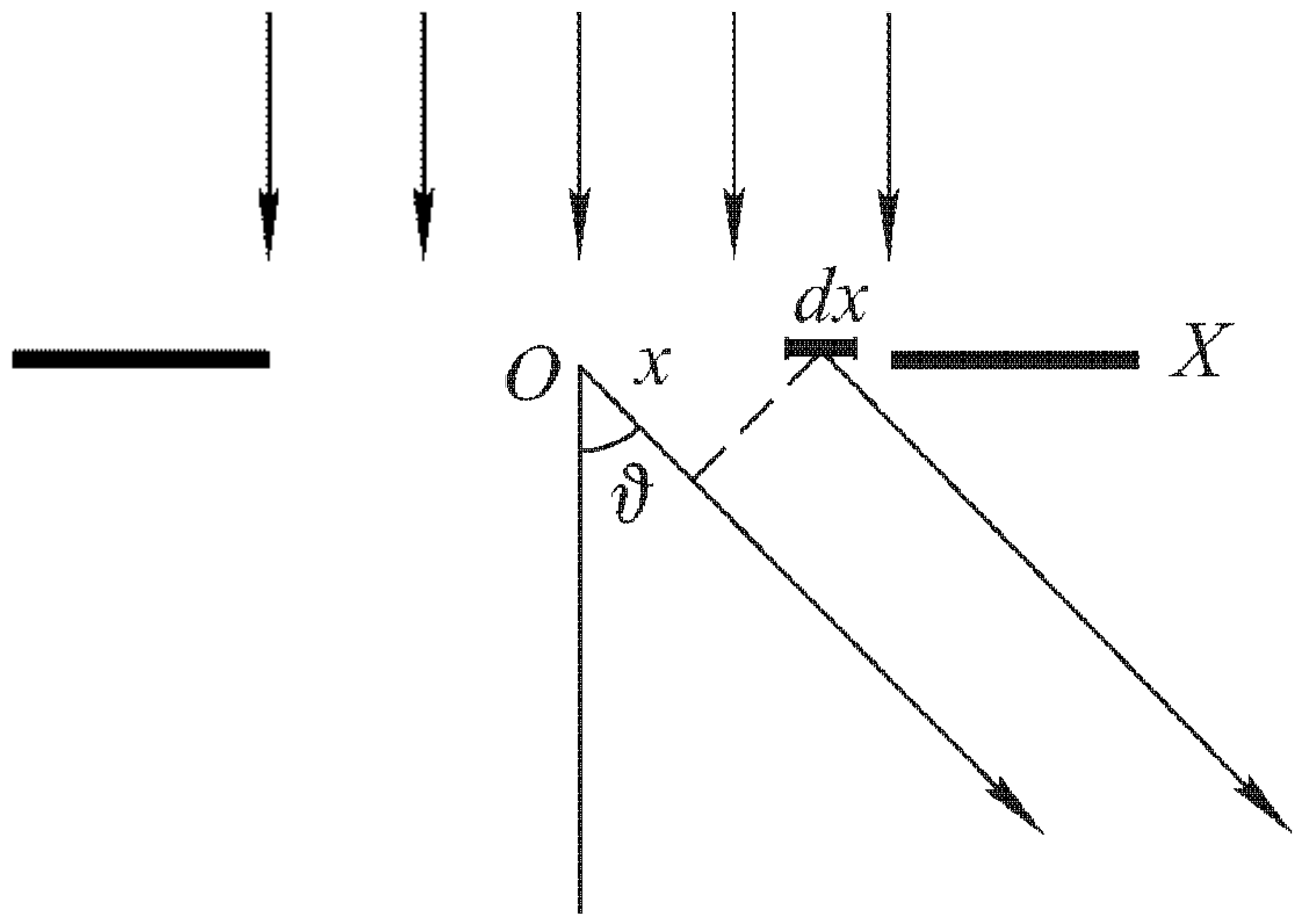


Рис. 174

Здесь опущены все множители, не влияющие на относительное распределение волнового поля по направлениям. Вычислив интеграл, получим

$$E = b\,\frac{\sin\alpha}{\alpha}, \tag{44.1}$$

где введено обозначение

$$\alpha = \frac{kb\sin\vartheta}{2} = \frac{\pi b\sin\vartheta}{\lambda}. \tag{44.2}$$

Отсюда для распределения интенсивности света по направлениям найдем

$$I = I_0\left(\frac{\sin\alpha}{\alpha}\right)^2, \tag{44.3}$$

где I_0 — интенсивность в направлении падающей волны. На рис. 175 представлены графики функций $\frac{\sin\alpha}{\alpha}$ (штриховая кривая) и $\left(\frac{\sin\alpha}{\alpha}\right)^2$ (сплошная кривая). Обе функции обращаются в максимум, равный единице, при $\alpha = 0$. При $\alpha = m\pi$, где $m = \pm1, \pm2, \ldots$, они равны нулю, т.е. в этих точках наблюдаются минимумы интенсивности. Между двумя соседними минимумами располагаются максимумы различных порядков. Их положения определяются трансцендентным уравнением $\alpha\cos\alpha - \sin\alpha = 0$. Практически можно считать, что максимумы располагаются посередине между соседними минимумами.

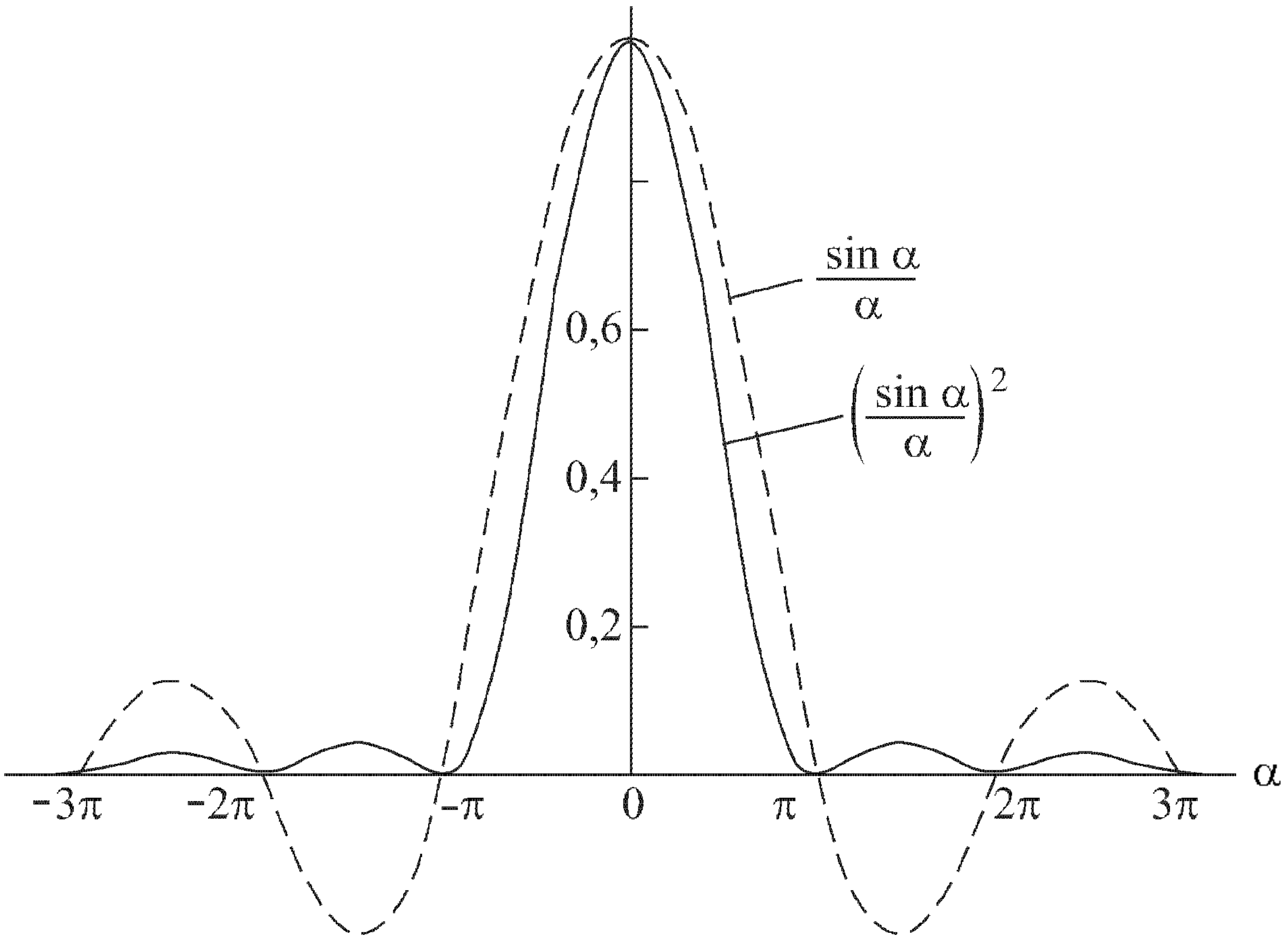


Рис. 175

2. Условие минимума $\alpha = m\pi$ можно также записать в виде

$$b\sin\vartheta = m\lambda. \tag{44.4}$$

Оно означает, что разность хода между волнами, исходящими от крайних точек щели, должна содержать целое число волн. Этот результат легко уяснить без всяких вычислений. Допустим сначала, что $b\sin\vartheta = \lambda$. Разобьем щель на две части одинаковой ширины. Тогда волны от этих частей придут в (удаленную) точку наблюдения в противоположных фазах и погасят друг друга при интерференции, — получится минимум интенсивности. Если $b\sin\vartheta = m\lambda$, то разбиение щели надо произвести на $2m$ частей одинаковой ширины. Волны от частей с нечетными номерами придут в фазах, противоположных фазам волн, пришедших от частей с четными номерами. А так как, ввиду симметрии, части щели, равноудаленные от противоположных краев ее, совершенно равноправны, то ясно, что эффект, вызванный нечетными частями, погасится при интерференции эффектом, вызванным четными частями.

Приведенное обоснование формулы (44.4) не только проще предыдущего, но, что особенно важно, и обладает большей общностью. Действительно, в предыдущем выводе щель должна предполагаться широкой ($b \gg \lambda$), чтобы можно было применять приближенный метод Френеля. Последний же вывод использует только свойства симметрии щели и применим не только при малых, но и при любых углах дифракции ϑ. Требуется только, чтобы было $b > \lambda$. В противном случае из условия (44.4) мы получили бы $\sin\vartheta > 1$, и все дифракционные минимумы с нулевой интенсивностью были бы невозможны.

3. Когда падающая волна плоская, то все дифрагированные лучи перпендикулярны к оси щели. Если на их пути поставить линзу, то каждый дифрагированный пучок параллельных лучей соберется в маленькое пятнышко. Такие пятнышки — максимумы и минимумы интенсивности — расположатся вдоль прямой, перпендикулярной к оси щели и лежащей в фокальной плоскости линзы. В этой плоскости и надо поместить экран для наблюдения.

Если в качестве источника света взять светящуюся линию или узкую освещаемую коллиматорную щель, параллельную щели, на которой происходит дифракция, то каждая точка источника даст на экране дифракционную картину, описанную выше. В результате наложения таких картин каждое дифракционное пятнышко вытянется в полоску. Образуется система дифракционных полос, показанная на рис. 176. Центральная полоса светлая и примерно вдвое шире остальных светлых и темных полос. Максимальный порядок минимума, который может наблюдаться, определяется условием $\sin\vartheta < 1$, т.е. $m < b/\lambda$, как это видно из формулы (44.4). Чем шире щель, тем ярче картина, тем у́же дифракционные полосы, а число самих полос больше. При сужении щели картина расширяется, а ее яркость уменьшается. Когда $b = \lambda$, минимумы первого порядка получаются при $\vartheta = \pi/2$ и исчезают при дальнейшем сужении щели.

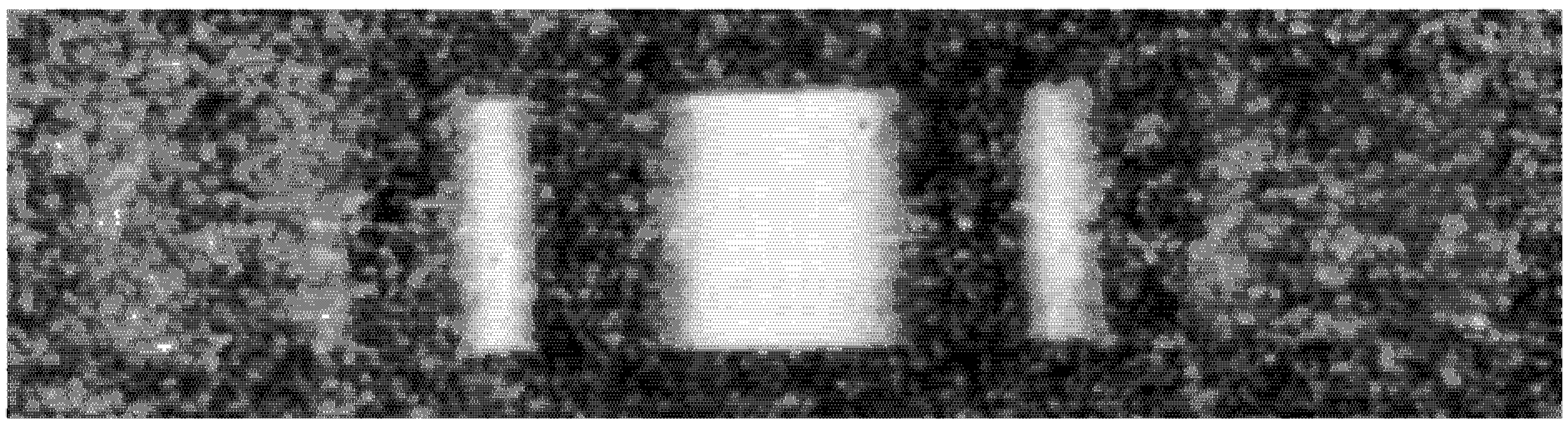


Рис. 176

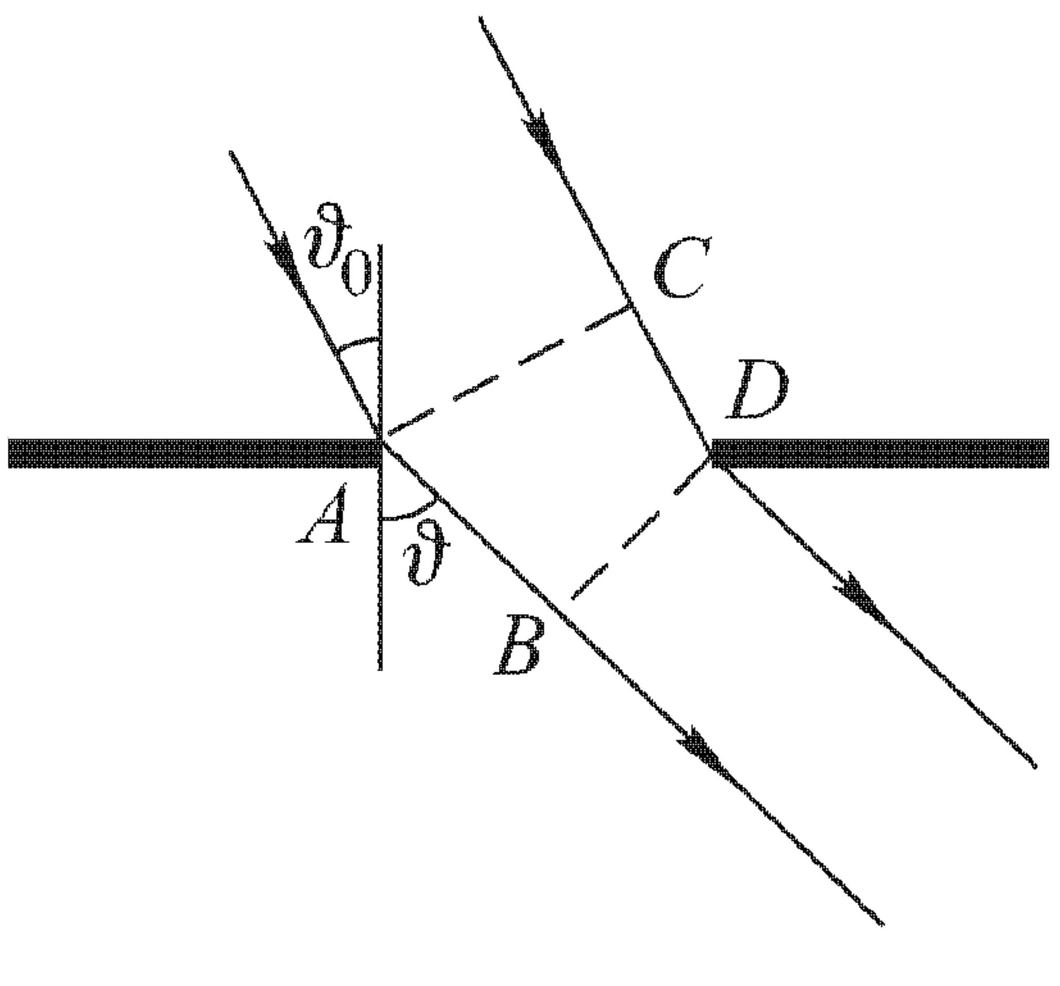


Рис. 177

4. При наклонном падении света под углом ϑ_0 (рис. 177) разность хода между крайними интерферирующими лучами $AB - CD$ составляет $b(\sin\vartheta - \sin\vartheta_0)$. Поэтому условие дифракционного минимума (44.4) переходит в

$$b\,(\sin\vartheta - \sin\vartheta_0) = m\lambda. \tag{44.5}$$

Если углы ϑ и ϑ_0 малы, то $b(\vartheta - \vartheta_0) \approx m\lambda$. Основная доля света сосредоточена в центральной дифракционной полосе, т.е. между минимумами первого и минус первого порядков. Этим можно воспользоваться для оценки дифракционной расходимости световых пучков, выделяемых тем или иным способом, например в результате прохождения плоской волны через диафрагму. Световые лучи, прошедшие через диафрагму, отклоняются от своего исходного направления на угол

$$\delta\vartheta \sim \lambda/D, \tag{44.6}$$

где D — поперечное сечение пучка (в направлении, где оно минимально). Такое дифракционное уширение пучка обусловлено волновой природой света и принципиально не может быть устранено при заданной ширине пучка. Поэтому не существует строго параллельных пучков лучей. Это — идеализация, предполагающая, что поперечное сечение пучка бесконечно велико. Если пучок проходит путь l, то на этом пути он претерпевает дифракционное уширение $\sim l\,\delta\vartheta = l\lambda/D$. Таким уширением можно пренебрегать только тогда, когда оно мало по сравнению с шириной самого пучка, т.е. когда $l\lambda/D \ll D$, или $l \ll D^2/\lambda$. Только на таких расстояниях пучок может рассматриваться как луч геометрической оптики (см. § 41, п. 3).

5. Допустим, что источником света является бесконечно удаленная светящаяся точка или линия, параллельная длине щели. Угол ϑ на дифракционный минимум m-го порядка определяется условием (44.5). Ближайший к нему максимум сдвинут на угол $\delta\vartheta$, причем

$$b\,[\sin(\vartheta + \delta\vartheta) - \sin\vartheta_0] = m\lambda + \frac{\lambda}{2}.$$

Вычитая отсюда (44.5) и заменяя разности дифференциалами, получим $b\cos\vartheta \cdot \delta\vartheta = \lambda/2$. Пусть другой такой же (некогерентный) источник света сдвинут относительно первого как раз на угол $\delta\vartheta$. Тогда дифракционные максимумы от одного из этих источников наложатся на минимумы от другого. В результате дифракционные полосы пропадут. Угловое расстояние $\delta\vartheta$ между точечными (или линейными) источниками, когда это произойдет, определяется формулой

$$\delta\vartheta = \lambda/2b \tag{44.7}$$

(при условии, что угол ϑ мал).

Допустим теперь, что источник света протяженный и имеет форму равномерно светящейся полосы с угловой шириной $\delta\Theta$. Разобьем ее на две одинаковые части. Каждую часть в свою очередь разобьем на бесконечно узкие полоски одинаковой ширины. Скомбинируем их в пары полосок, сдвинутых относительно друг друга на угол $\delta\Theta/2$. Если $\delta\Theta/2 = \delta\vartheta$, то, очевидно, дифракционных полос не получится. Таким образом, условие исчезновения дифракционных полос от протяженного источника рассматриваемой формы будет

$$\delta\Theta = \lambda/b. \tag{44.8}$$

От угловых размеров $\delta\Theta$ источника легко перейти к его линейным поперечным размерам δl. Если источник помещен в главном фокусе коллиматорной линзы с фокусным расстоянием f, то $\delta l = f\,\delta\Theta$, т.е.

$$\delta l = f\frac{\lambda}{b}. \tag{44.9}$$

6. Когда ширина щели становится меньше или порядка длины волны, приближенный метод Френеля, которым мы пользовались выше, становится неприменимым. Тогда волновое поле в плоскости щели уже нельзя отождествлять с неискаженным полем падающей волны, как это делается в методе Френеля. Задачу надо решать математически строго с использованием уравнений Максвелла и соответствующих им граничных условий.

Впервые такой метод был осуществлен в 1896 г. Зоммерфельдом (1868–1951) в задаче о дифракции плоской волны на прямолинейном крае экрана. Зоммерфельд рассмотрел идеально проводящий (а потому непрозрачный) экран, толщина которого пренебрежимо мала по сравнению с длиной волны. Хотя в оптике такой случай и невозможно осуществить, решение Зоммерфельда имеет большое значение, так как оно позволяет судить о точности и границах применимости приближенных методов. В 1897 г. Рэлей решил задачу о дифракции на узкой щели ($b \ll \lambda$) в бесконечно тонком идеально проводящем экране. В курсе общей физики нет возможности приводить эти решения¹). Сравним только их результаты с тем, что дает простой метод Френеля, чтобы составить более конкретное представление о границах применимости этого метода.

Предположим, что электрический вектор падающей волны параллелен щели. Как показывает строгий расчет, ход амплитуды дифрагированной волны качественно сохраняется, но выражается через функции Бесселя. Амплитудная кривая с ростом ϑ спадает круче прежней функции $(\sin\alpha)/\alpha$. В максимуме значение амплитуды в $4\lambda/(b\pi^2)$ раз меньше значения, определяемого формулой (44.1). Так, при $b = {}^1\!/_{10}\lambda$, амплитуда уменьшается в четыре раза²). Расхождение с приближенной теорией уменьшается при дальнейшем увеличении b.

---

¹) Интересующихся отсылаем к книге: *А. Зоммерфельд.* Оптика. — М.: ИЛ, 1953.

²) Если электрический вектор падающей волны перпендикулярен к длине щели, то амплитуда уменьшается слабее. Это легко понять, обратившись к соответствующему опыту Герца с проволочной решеткой (см. т. III, § 142, п. 7).

Численные расчеты, выполненные Морзе и Рубинштейном (1938), показали, что при ширине щели около λ или больше результаты, полученные по приближенному методу Френеля, могут считаться достаточно удовлетворительными. Таким образом, даже в случае наиболее тонких современных дифракционных решеток (при щелях порядка 1000–2000 нм) приближенный метод Френеля не ведет еще к заметным ошибкам.

§ 45. Дифракция Фраунгофера на отверстиях

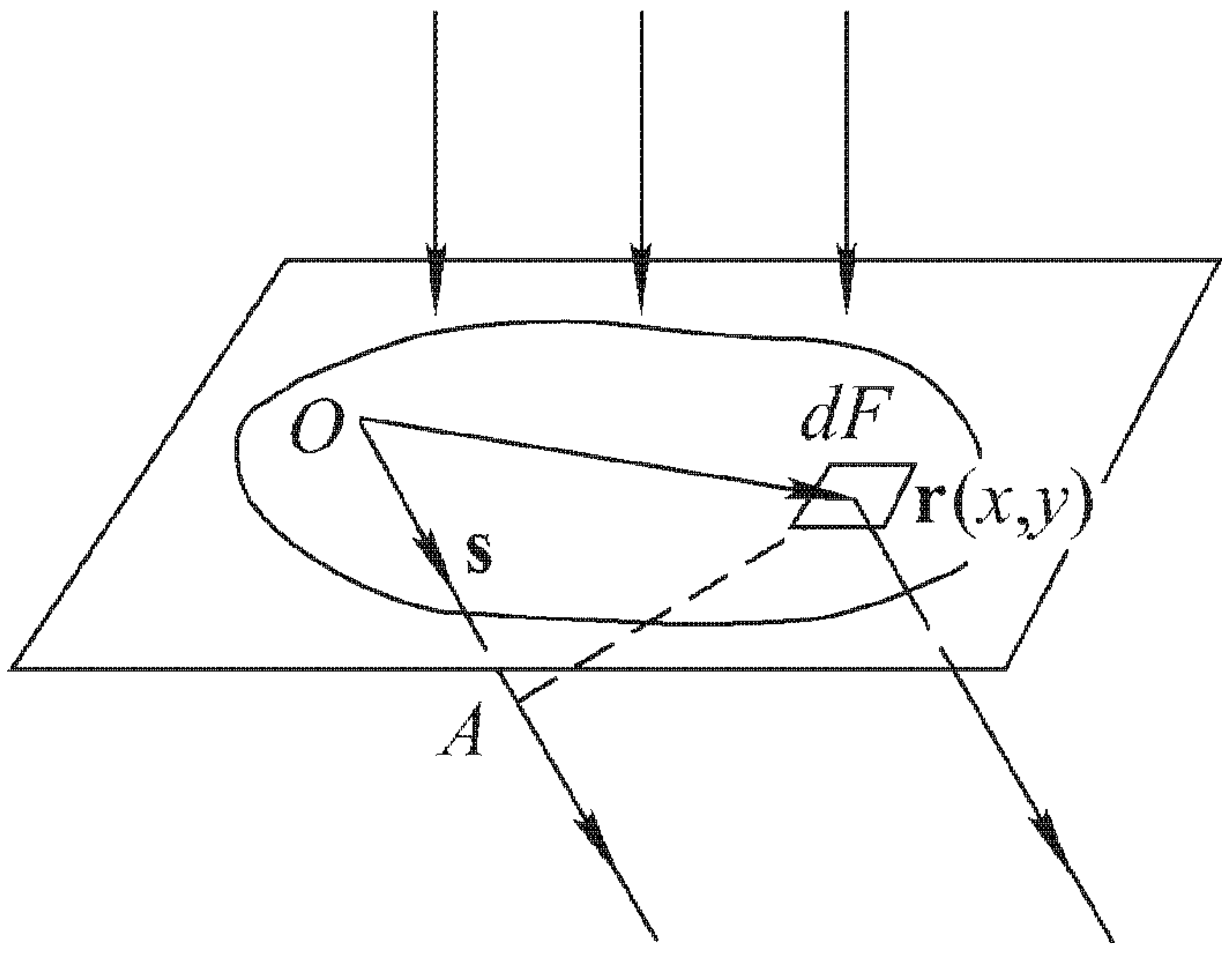


Рис. 178

1. Допустим, что свет падает перпендикулярно к плоскости непрозрачного экрана с отверстиями. (Обобщение на случай наклонного падения не встречает никаких затруднений.) Координатную плоскость XY совместим с плоскостью экрана, через dF обозначим элемент площади в этой плоскости. Направление дифрагированного света будем характеризовать единичным вектором $\mathbf{s}$ (рис. 178). Разность хода между лучами, вышедшими в этом направлении из элемента площади dF и из начала координат O, т.е. длина отрезка OA, равна $(\mathbf{rs})$, где $\mathbf{r}(x, y)$ — радиус-вектор элемента dF. Соответствующая разность фаз будет $k(\mathbf{rs})$. Результирующее поле в фраунгоферовой дифракционной картине представится интегралом

$$E = \int e^{ik(\mathbf{sr})}\, dF, \tag{45.1}$$

распространенным по всем отверстиям.

В случае прямоугольного отверстия удобно перейти к прямоугольным координатам, предполагая, что координатные оси параллельны сторонам отверстия. Если a и b — длины этих сторон, то

$$E = \int\limits_{-a/2}^{+a/2} \int\limits_{-b/2}^{+b/2} e^{ik(s_x x + s_y y)}\, dx\, dy. \tag{45.2}$$

Это — в точности такие же интегралы, которые встречались нам при рассмотрении дифракции на щели. Выполнив интегрирование, получим

$$E = ab\,\frac{\sin\alpha}{\alpha}\,\frac{\sin\beta}{\beta}, \tag{45.3}$$

где

$$\alpha = \frac{1}{2}\,kas_x = \frac{\pi a s_x}{\lambda}, \qquad \beta = \frac{1}{2}\,kbs_y = \frac{\pi b s_y}{\lambda}. \qquad (45.4)$$

Заменив в одном из интегралов пределы интегрирования бесконечными, получим предельный случай бесконечно длинной щели. Интенсивность определяется формулой

$$I = I_0 \left(\frac{\sin\alpha}{\alpha}\right)^2 \left(\frac{\sin\beta}{\beta}\right)^2. \qquad (45.5)$$

Дифракционную картину можно получить, если наложить друг на друга две взаимно перпендикулярные дифракционные картины, одна из которых получена при дифракции на щели ширины a, а другая — на щели ширины b (рис. 179). Картина вытянута в направлении более короткой стороны прямоугольного отверстия.

Случай круглого отверстия на практике представляет большой интерес, так как все оправы линз и объективов имеют обычно круглую форму. В этом случае при вычислении интеграла (45.1) естественно перейти к полярным координатам. При малых углах дифракции интеграл выражается через бесселеву функцию первого порядка $J_1(\alpha)$, где $\alpha = kR\vartheta = 2\pi R\vartheta/\lambda$ (R — радиус отверстия, ϑ — угол дифракции). Опуская вычисления, приведем окончательные результаты. Дифракционная картина, естественно, имеет вид концентрических светлых и темных колец (рис. 180). Центр картины светлый, так как в него все вторичные волны приходят в оди-

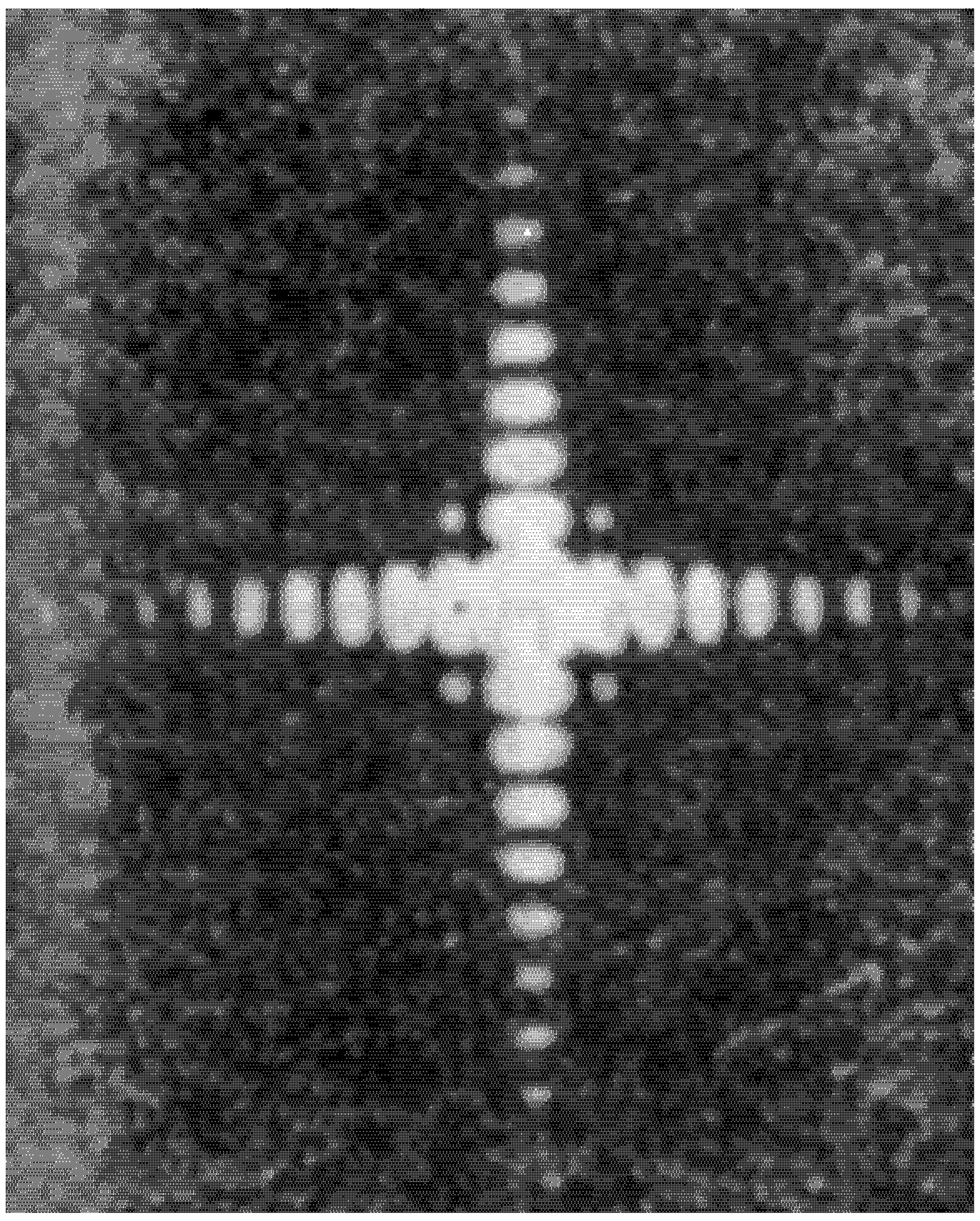


Рис. 179

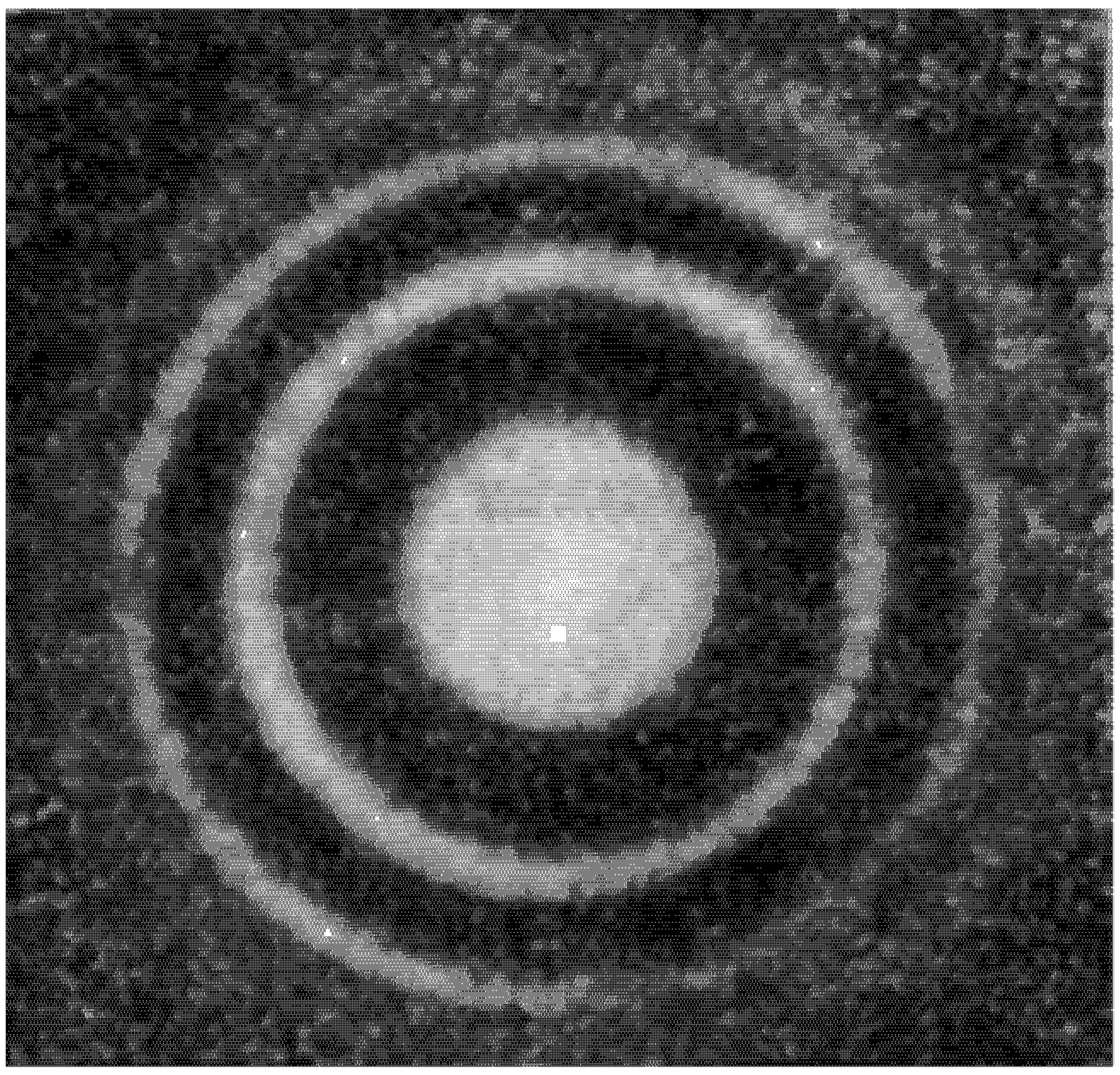


Рис. 180

наковых фазах. Распределение амплитуд (штриховая кривая) и интенсивностей (сплошная кривая) в зависимости от угла дифракции ϑ (или, что то же, расстояния от центра картины) приведено на рис. 181. Соответствующие кривые мало отличаются от кривых рис. 175 не только качественно, но и количественно. Приближенно угловые радиусы темных колец определяются формулой

$$\vartheta_m = \left[0{,}61 + \frac{m-1}{2}\right]\frac{\lambda}{R}. \tag{45.6}$$

Более точные данные приведены в табл. 5. Из нее видно, что около 98 % света приходится на центральный максимум. Если исключить центральный максимум, то остальные темные и светлые кольца практически равноотстоящие.

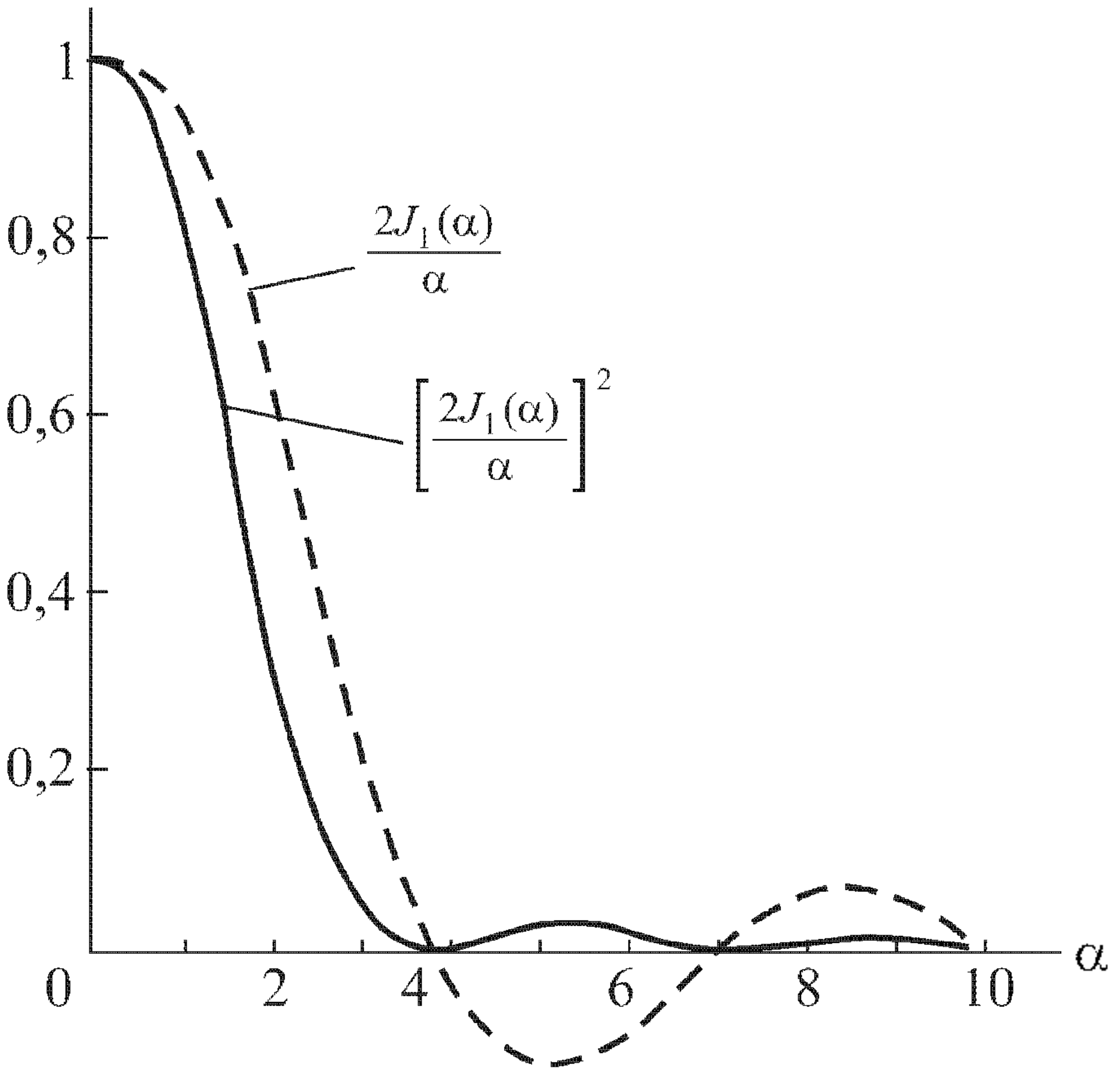


Рис. 181

Таблица 5

Минимумы	Максимумы	Интенсивность в максимумах
$\vartheta_1 = 0{,}61$	$\vartheta'_1 = 0$	1
$\vartheta_2 = 1{,}12$	$\vartheta'_2 = 0{,}81$	0,0175
$\vartheta_3 = 1{,}62$	$\vartheta'_3 = 1{,}33$	0,0042
$\vartheta_4 = 2{,}12$	$\vartheta'_4 = 1{,}85$	0,0016

2. Таким же путем, по крайней мере численно, можно рассчитать дифракционную картину Фраунгофера на отверстии любой формы. При решении подобных задач полезно руководствоваться *соображениями подобия*. Представим волновое поле интегралом вида (45.2), но распространенным по области F, занимаемой рассматриваемым отверстием. Введем новые координаты $x' = \mu x$, $y' = y$, где μ — постоянная. Область F плоскости XY преобразуется в область F' плоскости $X'Y'$. Она получается из

F равномерным растяжением в μ раз в направлении оси X. Интеграл (45.2) преобразуется в

$$E(\mathbf{s}) = \frac{1}{\mu} \iint\limits_{F'} e^{i(s'_x x' + s'_y y')}\, dx'\, dy',$$

где единичный вектор $\mathbf{s}'$ определяется своими проекциями $s'_x = s_x/\mu$, $s'_y = s_y$. (Постоянная μ должна быть такой, чтобы $s'^2_x + s'^2_y < 1$.) Но последний интеграл представляет волновое поле $E'(\mathbf{s}')$ (в штрихованной системе координат) в направлении единичного вектора $\mathbf{s}'$ при фраунгоферовой дифракции на отверстии F', а потому

$$E'(\mathbf{s}') = \mu E(\mathbf{s}). \tag{45.7}$$

Таким образом, по известной дифракционной картине на каком-либо отверстии можно без новых вычислений получить новые дифракционные картины. Для этого надо отверстие равномерно вытянуть (сжать) в каком-либо направлении. Тогда, как видно из формул $x' = \mu x$, $s'_x = s_x/\mu$, дифракционная картина сожмется (вытянется) в том же направлении. Так, при растяжении круглого отверстия оно переходит в эллиптическое, а дифракционные кольца сжимаются, также принимая эллиптическую форму. Конечно, отверстие можно вытянуть или сжать и вдоль каких-либо двух направлений.

3. Рассмотрим теперь случай, когда в экране имеется большое число N одинаковых и одинаково ориентированных отверстий. Волновое поле в бесконечности представится суммой $E = \sum E_i$, где E_i — поле, которое возникло бы при дифракции при наличии одного только i-го отверстия. Для интенсивности получим

$$I = \overline{EE^*} = \sum I_i + \sum_{i \neq j}\sum \overline{E_i E_j^*}.$$

При фраунгоферовой дифракции распределение интенсивности в дифракционной картине определяется только *направлением лучей*, а не положением световых пучков. При боковом смещении последних интенсивность не меняется. *Распределение интенсивности не изменится, если отверстие в плоскости экрана сместить в сторону без изменения его ориентации.* Поэтому в последней сумме все интенсивности одинаковы: $I_i = I_1$.

Рассмотрим особо два случая: 1) отверстия расположены хаотически; 2) отверстия расположены «правильно», в определенном порядке. В первом случае среди членов двойной суммы с $i \neq j$ в среднем найдется столько же положительных членов, сколько и отрицательных. При сложении таких членов в среднем получится нуль. Поэтому $I = NI_1$. Получается такая же дифракционная картина, что и от одного отверстия, но усиленная по интенсивности в N раз. Интенсивности отдельных картин арифметически складываются, но сами картины не интерферируют между собой. Во втором случае, напротив, члены с $i \neq j$ не компенсируются. Они могут интерферировать и существенно влиять на дифракционную картину. Этот случай будет рассмотрен в следующем параграфе на примере дифракционной решетки.

Первый случай легко демонстрировать на стеклянной пластинке, запыленной спорами ликоподия. Последние имеют форму шариков практически одинаковых размеров. При освещении пластинки параллельным пучком лучей на удаленном экране появляется дифракционная картина, состоящая из концентрических колец. При освещении белым светом внешние края колец окрашены в красный, а внутренние — в фиолетовый цвет. Это указывает на дифракционную природу явления, так как длины волн красных лучей больше, чем фиолетовых.

4. Описанное явление наблюдается в природе в виде *венцов*. Так называются светлые туманные кольца на небесном своде вокруг Солнца или Луны. Иногда венцы наблюдаются вокруг ярких звезд или планет, а также вокруг земных источников света. Венцы возникают в результате дифракции света на водяных капельках (или кристалликах льда), когда перед светилом проходит полупрозрачное облако (чаще всего высококучевое) или туман. Угловые радиусы венцов обычно не превосходят $5\,^\circ$. Дифракционная природа колец в явлении венцов подтверждается тем, что наружные края колец имеют красноватый цвет, а внутренние — синеватый. При наличии в атмосфере капель всевозможных размеров кольца венцов налагаются друг на друга и образуют общее белое сияние вокруг диска светила, называемое в случае Солнца *околосолнечным ореолом*.

От венцов следует отличать *гало*. Так называется группа оптических явлений в атмосфере, возникающая при преломлении или отражении лучей Солнца или Луны на плавающих в воздухе кристалликах льда, образующих перистые облака. Угловые радиусы гало значительно больше угловых радиусов венцов и составляют 22° или 46°. Наружный край колец гало имеет синеватую, а. внутренний — красноватую окраску. Это указывает на рефракционную природу явления (дисперсия). В отличие от венцов, угловые размеры которых могут меняться с изменением радиусов водяных капелек, угловые размеры колец гало строго постоянны, так как они зависят только от углов между гранями кристалликов, которые при изменении размеров последних остаются неизменными.

§ 46. Дифракционная решетка

1. *Дифракционная решетка* — важнейший спектральный прибор, предназначенный для *разложения света в спектр и измерения длин волн.* Она представляет собой плоскую стеклянную или металлическую поверхность, на которой делительной машиной нарезано очень много (до сотен тысяч) прямых равноотстоящих штрихов. На стеклянных решетках наблюдения можно производить как в *проходящем*, так и в *отраженном свете*, на металлических — только в *отраженном.* Применяются *вогнутые металлические решетки*, в которых штрихи наносятся на вогнутой сферической поверхности.

Сначала рассмотрим простейшую идеализированную решетку, состоящую из одинаковых равноотстоящих параллельных щелей, сделанных в непрозрачном экране. Ширину щели обозначим через b, ширину непрозрачной части экрана между двумя соседними щелями — через a. Величина $d = a+b$ называется *периодом решетки.* В решетке осуществляется *многолучевая интерференция* когерентных дифрагированных пучков света, исходящих от щелей решетки при ее освещении. Дифракционная картина наблюдается по методу Фраунгофера, т.е. либо на бесконечно удаленном экране, либо в фокальной плоскости линзы, поставленной на пути дифрагированного света.

Пусть на решетку перпендикулярно к ее поверхности падает плоская монохроматическая волна (рис. 182). Разность хода между вторичными волнами, исходящими из соседних щелей решетки, будет $d\sin\vartheta$, а разность фаз $\delta = kd\sin\vartheta = \frac{2\pi}{\lambda}d\sin\vartheta$, где ϑ — угол дифракции. Обозначим через E_1 поле в точке наблюдения, излучаемое первой щелью. Оно определяется формулой

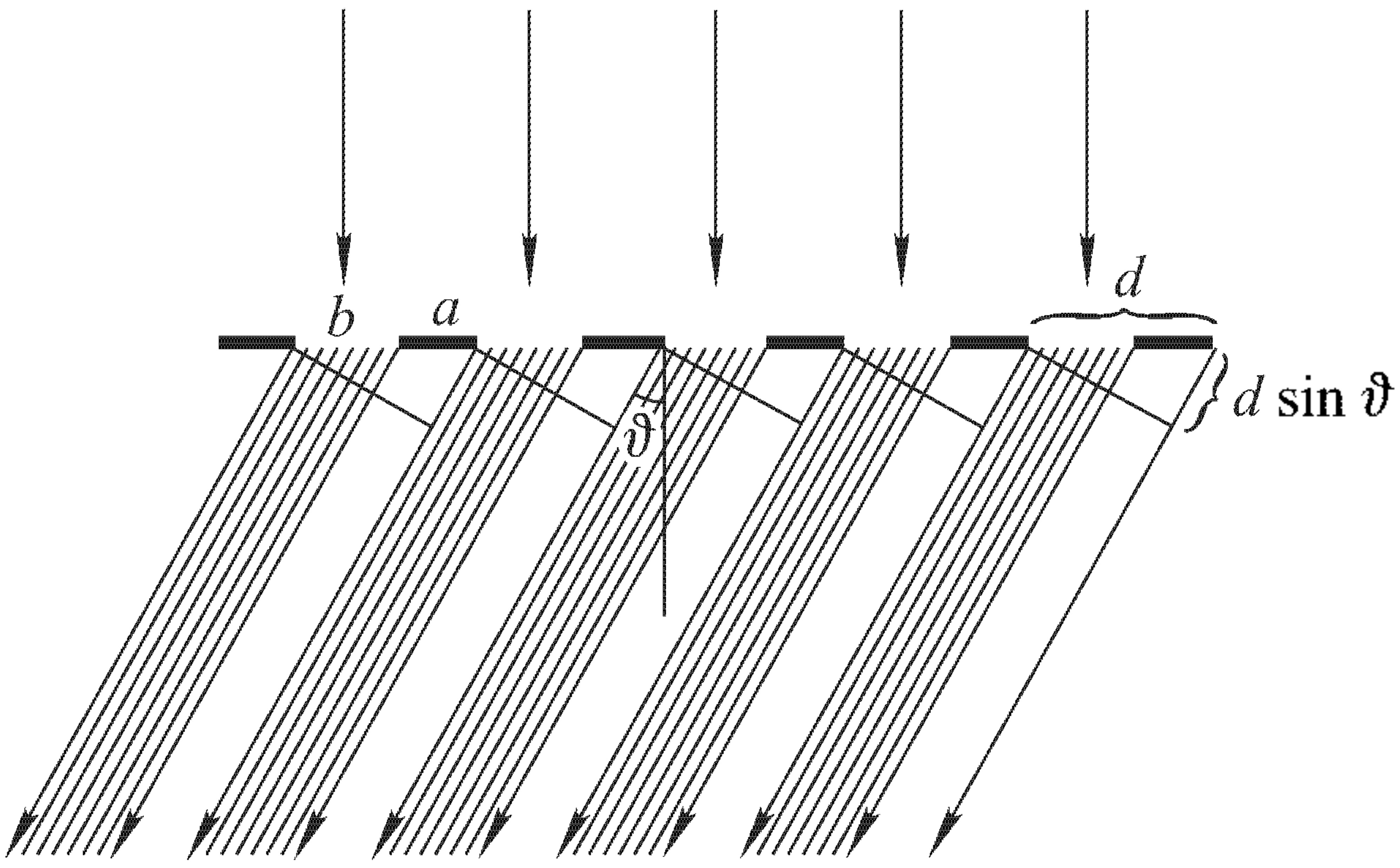


Рис. 182

$E_1 = b\dfrac{\sin\alpha}{\alpha}$. Поля, излучаемые остальными щелями, представляются выражениями:

$$E_2 = E_1 e^{-i\delta}, \quad E_3 = E_1 e^{-2i\delta}, \ldots, E_N = E_1 e^{-i(N-1)\delta},$$

где N — общее число щелей. Полное поле, излучаемое всеми щелями, представляется суммой

$$E = E_1[1 + e^{-i\delta} + e^{-2i\delta} + \ldots + e^{-i(N-1)\delta}] = E_1\frac{1 - e^{-iN\delta}}{1 - e^{-i\delta}},$$

откуда

$$E = E_1\frac{\sin(N\delta/2)}{\sin(\delta/2)}e^{-i(N-1)\delta/2},$$

$$A = A_1\frac{\sin(N\delta/2)}{\sin(\delta/2)}, \tag{46.1}$$

где A_1 — вещественная амплитуда волны от одной щели, A — от всей решетки. Для интенсивностей получаем

$$I = I_1\left[\frac{\sin(N\delta/2)}{\sin(\delta/2)}\right]^2. \tag{46.2}$$

Формулы (46.1) и (46.2) — основные в теории дифракционной решетки.

2. Когда $\vartheta = 0$, то $\delta = 0$. В этом случае выражения (46.1) и (46.2) принимают неопределенный вид $0/0$. Раскрыв неопределенность, получим

$$A_{\text{гл}} = NA_1, \quad I_{\text{гл}} = N^2 I_1. \tag{46.3}$$

Тот же результат получается в случае $\delta/2 = m\pi$, т.е. при

$$d\sin\vartheta = m\lambda \qquad (m = 0, \pm1, \pm2, \ldots). \tag{46.4}$$

В направлениях, определяемых этим условием, получаются максимумы, интенсивность которых в N^2 раз превосходит интенсивность волны от одной щели в том же направлении. Они называются *главными максимумами*. Целое число m называют *порядком главного максимума* или *порядком спектра*. Условие (46.4) определяет направления, в которых излучения от всех щелей решетки приходят в точку наблюдения в одинаковых фазах, а потому усиливают друг друга. В таких направлениях при отдельных значениях m могут и не возникнуть максимумы. Это будет, когда $I_1 = 0$, т.е. в направлениях на дифракционные минимумы от одной щели. Например, если $a = b$, то все главные максимумы четных порядков не появятся. Действительно, условие появления главного максимума порядка $2n$ имеет вид $d\sin\vartheta = 2n\lambda$. При $d = 2b$ оно переходит в $b\sin\vartheta = n\lambda$, т.е. в условие *дифракционного минимума* на щели. Таким образом, в рассматриваемом направлении ни одна щель, а потому и решетка в целом не излучают.

3. Выражения (46.1) и (46.2) обращаются в нуль, если $\sin(N\delta/2) = 0$, но $\sin(\delta/2) \neq 0$, т.е. при $N\delta/2 = (Nm+p)\pi$, или

$$d\sin\vartheta = \left(m + \frac{p}{N}\right)\lambda \qquad (p = 1, 2, \ldots, N-1). \tag{46.5}$$

В соответствующих направлениях получаются дифракционные минимумы, в которых интенсивность света равна нулю. Между двумя соседними минимумами получается максимум. Такие максимумы называются *второстепенными* или *добавочными*. Между двумя соседними главными максимумами располагается $(N-1)$ минимумов и $(N-2)$ добавочных максимумов. На эти максимумы и минимумы накладываются минимумы, возникающие при дифракции от отдельной щели, в которых функция E_1 обращается в нуль.

Второстепенные максимумы находятся примерно посередине между соответствующими дифракционными минимумами. Величину δ, определяющую направление на какой-либо из второстепенных максимумов, можно поэтому вычислить по приближенной формуле

$$N\frac{\delta}{2} = \frac{(Nm+p) + (Nm+p+1)}{2}\,\pi,$$

или

$$\frac{\delta}{2} = \left(m + \frac{2p+1}{2N}\right)\pi. \tag{46.6}$$

Пользуясь этой формулой, найдем приближенное выражение для интенсивности второстепенных максимумов в окрестности соответствующего главного максимума (т.е. при малых p). В положениях второстепенных максимумов числители в формуле (46.2) равны единице. Если число щелей решетки N очень велико, а номер второстепенного максимума p невелик, то угол $\delta/2$ будет мал, и можно положить

$$\sin\frac{\delta}{2} = \pm\sin\frac{2p+1}{2N}\,\pi \approx \pm\frac{2p+1}{2N}\,\pi.$$

Это дает

$$I = \frac{I_1}{\pi^2}\left(\frac{2N}{2p+1}\right)^2 = \frac{4}{(2p+1)^2\pi^2}\,I_{\text{гл}}. \tag{46.7}$$

Таким образом, интенсивности главного максимума и ближайших к нему второстепенных максимумов находятся в отношениях

$$1 : \frac{4}{9\pi^2} : \frac{4}{25\pi^2} : \frac{4}{49\pi^2} : \ldots = 1 : 0{,}045 : 0{,}016 : 0{,}0083 : \ldots$$

Второстепенные максимумы слабы по сравнению с главными максимумами. При большом числе щелей они обычно не играют роли. Второстепенные максимумы создают более или менее равномерный слабый фон, на нем выступают узкие и резкие главные максимумы, в которых концентрируется практически весь дифрагированный свет. Распределение интенсивности в дифрагированном свете представлено на схематическом рис. 183 для

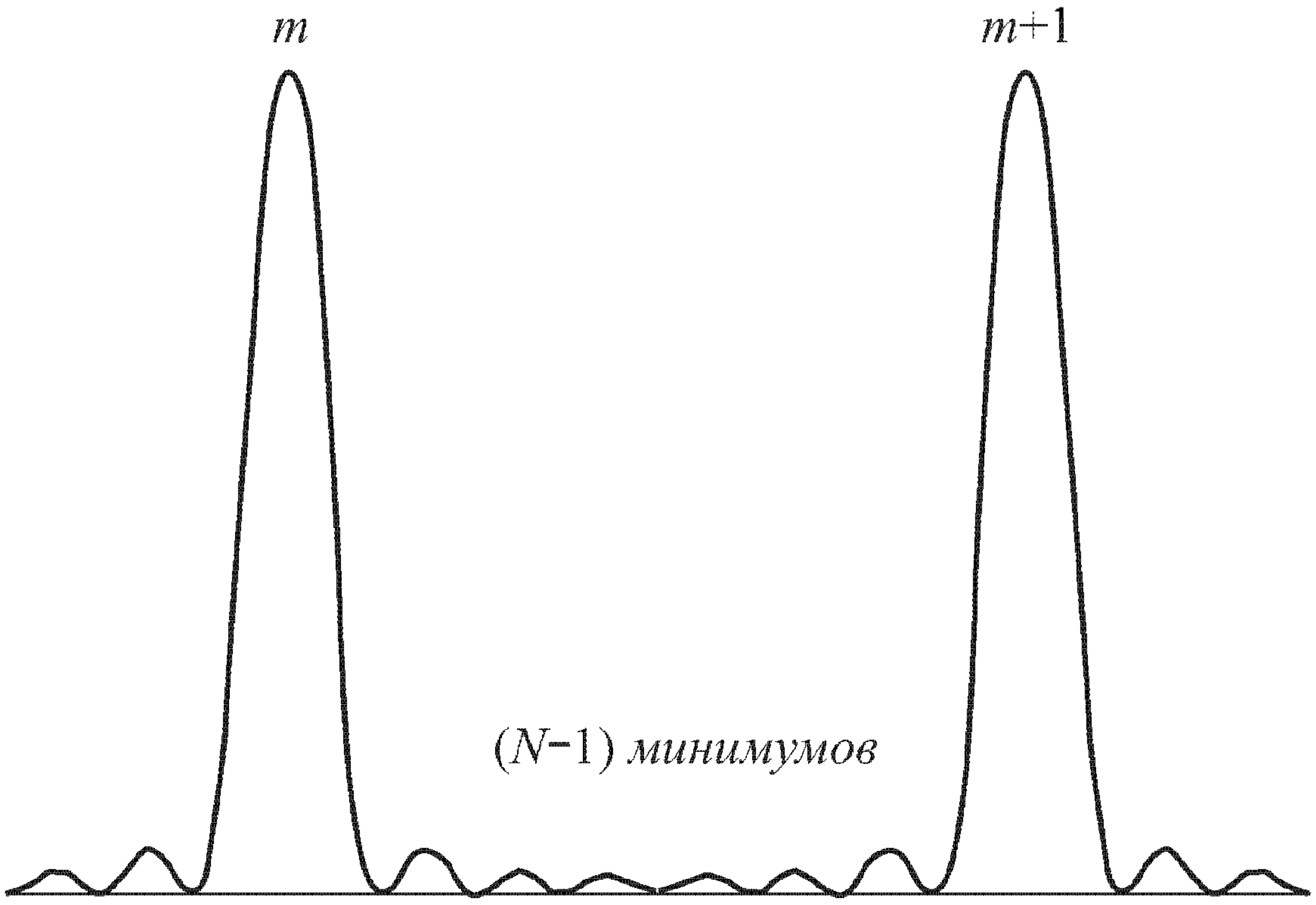


Рис. 183

$N = 8$. Величина множителя I_1, входящего в формулу (46.2), не указана. Между соседними главными максимумами находится $N - 2 = 6$ второстепенных максимумов. Относительные интенсивности главных и второстепенных максимумов представляются числами 100; 5,0; 2,25; 1,6; 1,6; 2,25; 5,0; 100.

Красивую демонстрацию можно получить, направив на решетку яркий узкий пучок света от лазера. При дифракции пучок расщепляется на много хорошо видимых в воздухе ярких пучков, веерообразно расходящихся от решетки. Падая на потолок и стены аудитории, эти пучки оставляют на них яркие светлые пятна (световые зайчики).

4. Происхождение второстепенных или добавочных максимумов и минимумов легко уяснить с помощью векторной диа-

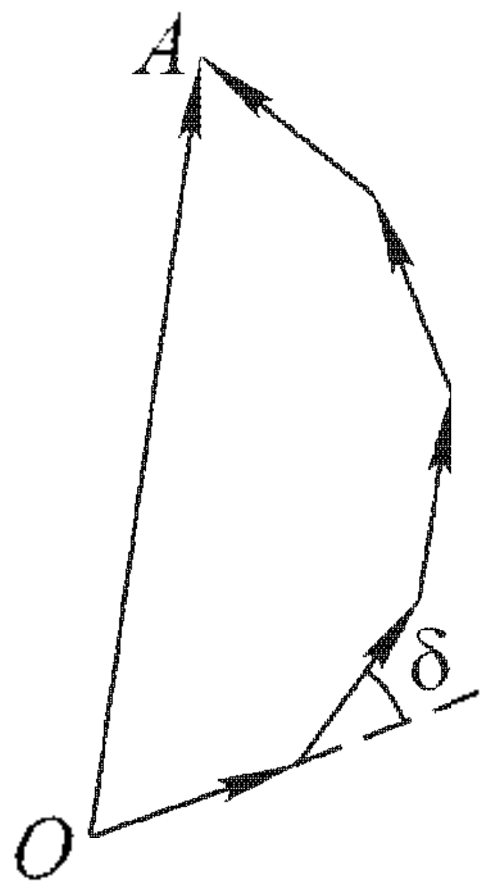


Рис. 184

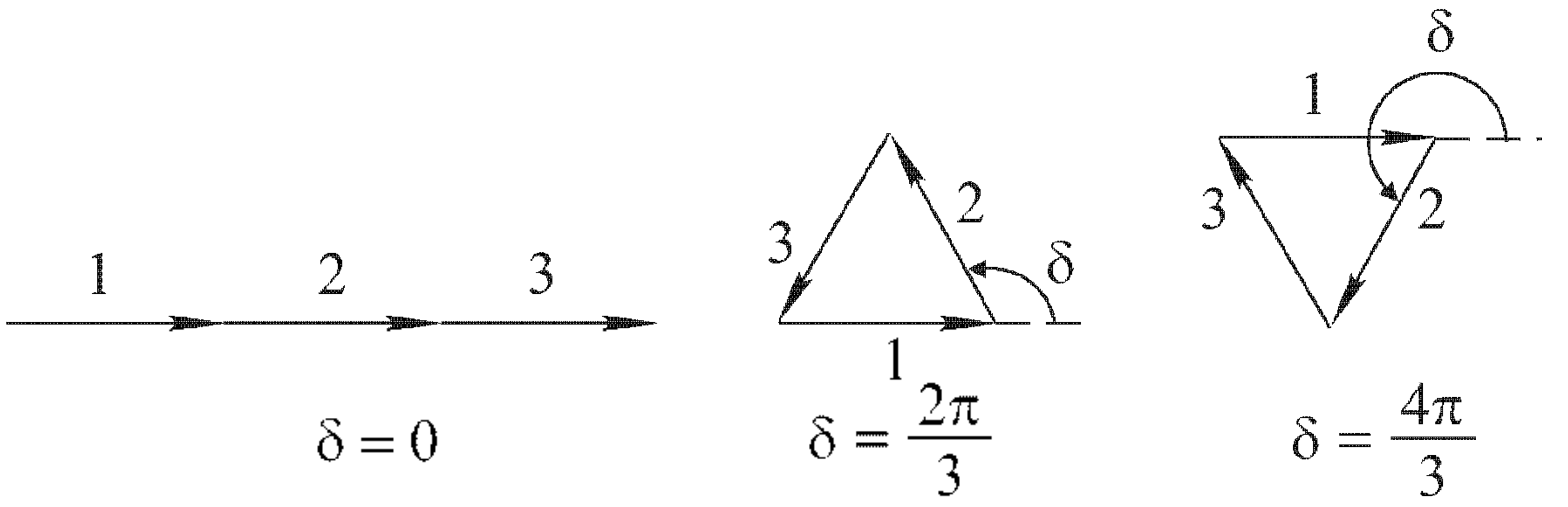


Рис. 185

граммы. На векторной диаграмме колебания от отдельных щелей решетки представляются векторами равной длины. Из-за сдвига фаз такие векторы повернуты один относительно другого на один и тот же угол $\delta = \frac{2\pi}{\lambda} d \sin\vartheta$ (рис. 184), образуя звенья правильной ломаной линии. Замыкающая этой ломаной OA изображает колебание, возбуждаемое всей решеткой. Если ломаная замкнутая, то $E = 0$, т.е. получается дифракционный минимум. Первый раз это произойдет при $\delta = 2\pi/N$, второй — при $\delta = 4\pi/N$ и т.д. Легко видеть, что таким путем получается условие дифракционного минимума (46.5). Все это проиллюстрировано на рис. 185 для $N = 3$ и на рис. 186 для $N = 4$.

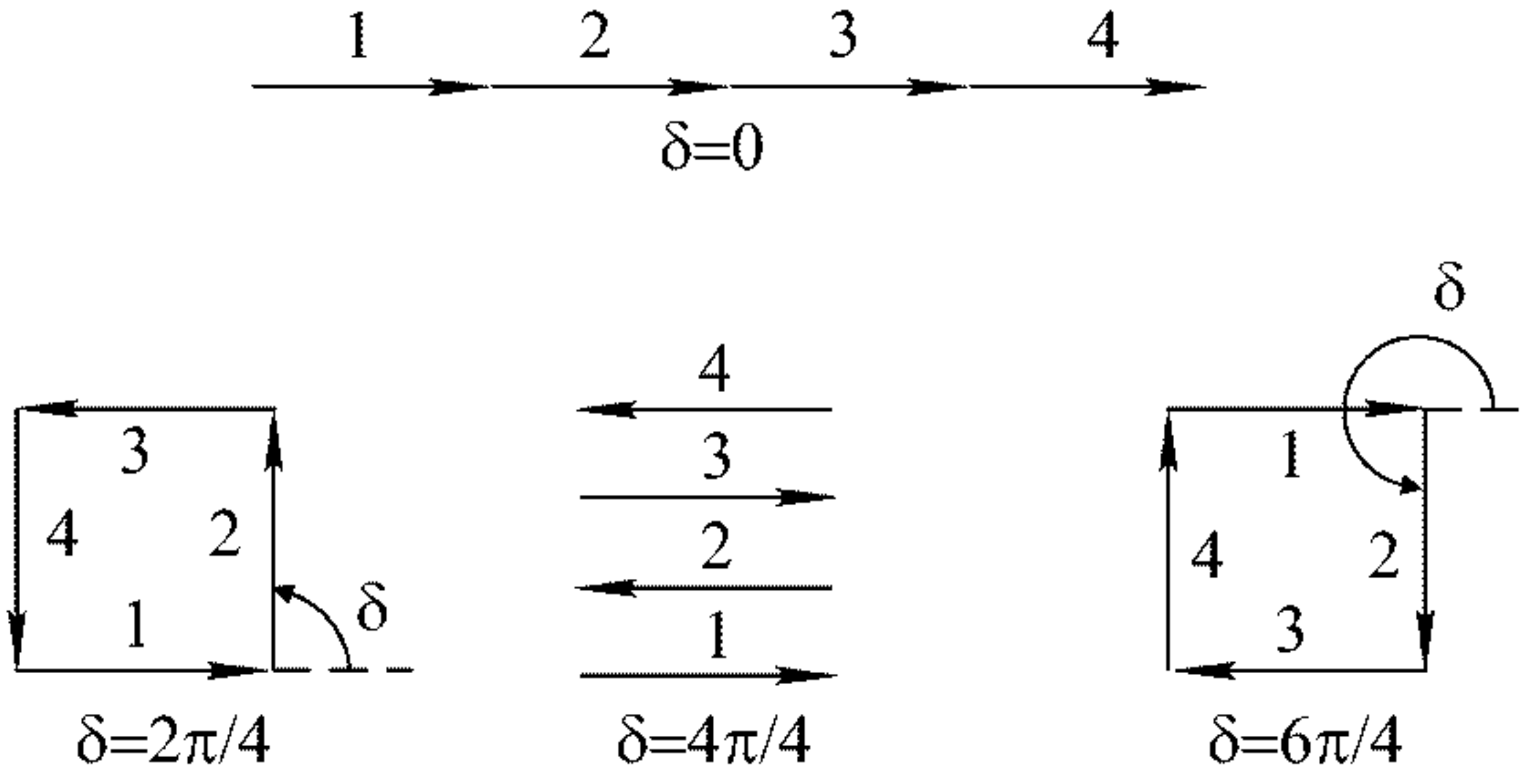


Рис. 186

5. Если волна падает на решетку наклонно под углом ϑ_0 (рис. 187), то разность хода между соседними пучками становится равной $AD - CB = d(\sin\vartheta - \sin\vartheta_0)$. Характер дифракционной картины в основном сохраняется. Положение главных максимумов определяется условием

$$d(\sin\vartheta - \sin\vartheta_0) = m\lambda, \quad (46.8)$$

а дифракционных минимумов — условием

$$d(\sin\vartheta - \sin\vartheta_0) = \left(m + \frac{p}{N}\right)\lambda \quad (p = 1, 2, \ldots, N-1). \quad (46.9)$$

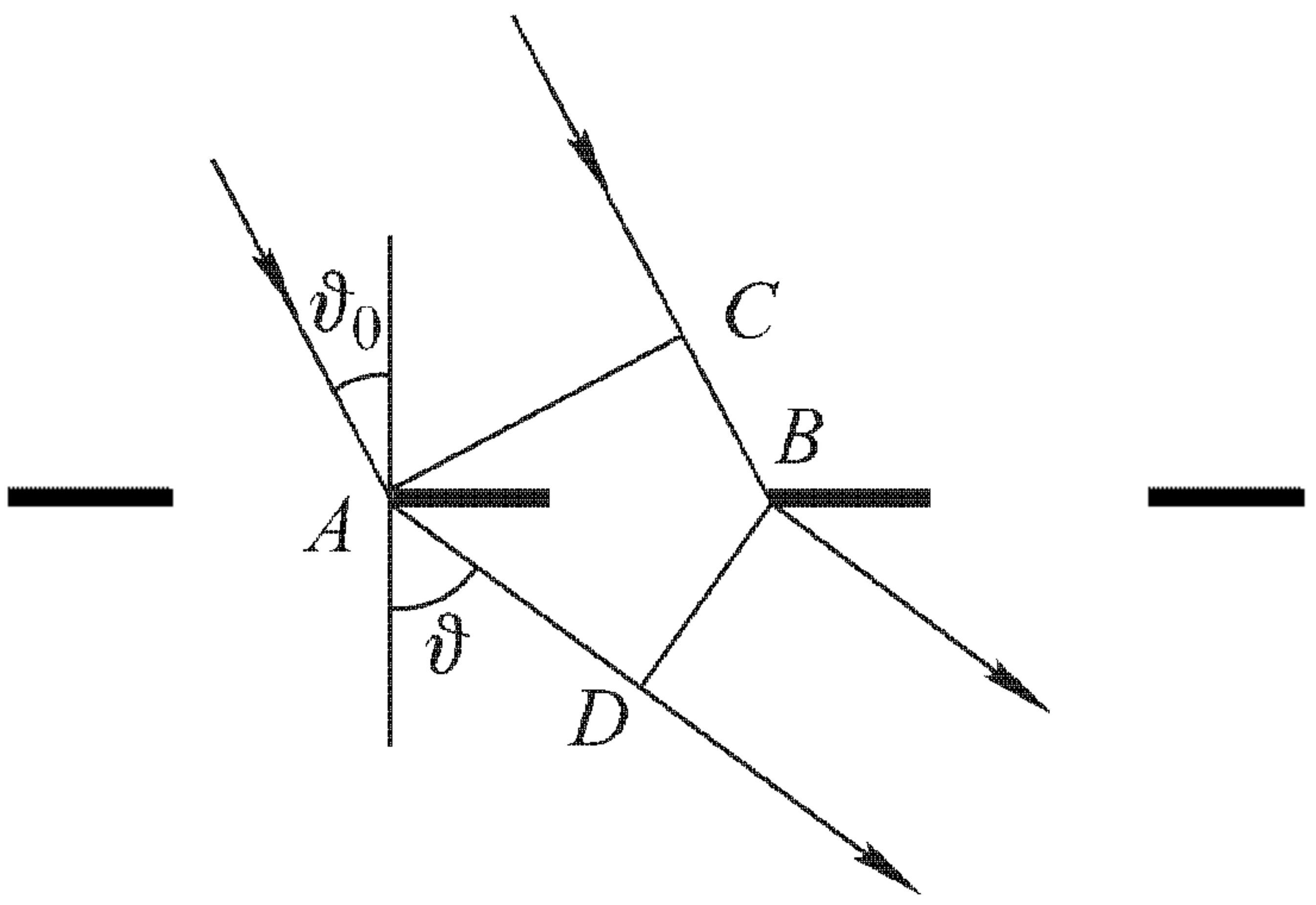


Рис. 187

Зная положение главных максимумов, можно вычислить длину волны по формуле (46.4) или (46.8). Решетка есть прибор для измерения именно *длины волны*, а не частоты колебаний, как это иногда неверно утверждают.

6. Форма штрихов, нанесенных на решетку, материал, из которого она изготовлена, и т.п. сказываются лишь на виде волны

E_1 *от отдельного штриха.* Положение же главных максимумов определяется исключительно периодом решетки, а дифракционных минимумов — кроме того еще числом штрихов N. При выводе формул (46.4), (46.8) и (46.9) из принципа Гюйгенса–Френеля нет необходимости применять приближенные методы. Из рассуждений п. 4 ясно, что существенна только *периодичность светового поля на выходе* (т.е. задней стороне) *решетки*, обусловленная в свою очередь периодичностью структуры последней. Ничего другого о поле на выходе решетки знать не надо. Поэтому перечисленные формулы *совершенно точны.*

Под дифракционной решеткой в широком смысле слова понимается всякая структура, обладающая *пространственной периодичностью.* Если свойства структуры периодически меняются только в одном направлении, то решетка называется *одномерной* или *линейной.* Если же периодичность решетки имеет место в двух или трех направлениях, то решетка называется соответственно *двух-* или *трехмерной.* В последнем случае ее называют также *пространственной.* В этом и следующих параграфах рассматриваются только линейные решетки.

Не обязательно, чтобы при прохождении через решетку менялась амплитуда волны. Существенно только, чтобы на выходе решетки периодически менялось *волновое поле в целом.* Можно различать два крайних идеализованных случая: 1) решетка вносит периодические изменения в амплитуду волны, не влияя на ее фазу; такая решетка называется *амплитудной*; 2) решетка вносит периодические изменения в фазу волны, но не влияет на ее амплитуду; такую решетку называют *фазовой.*

Всякая реальная решетка, строго говоря, не является чисто амплитудной или чисто фазовой. Она периодически меняет на выходе как амплитуду, так и фазу волнового поля. Приблизительно амплитудной решеткой является рассмотренная выше совокупность равноотстоящих щелей в непрозрачном экране (рис. 187). Приближением фазовой решетки может служить стеклянная пластинка, представленная на рис. 188. В обоих случаях период решетки должен быть велик по сравнению с длиной волны. Примером фазовой отражательной решетки может служить решетка, изготовленная С.М. Рытовым (р. 1908) и И.Л. Фабелинским (р. 1911). Она представляет собой равнобочную стеклянную призму с преломляющим углом 90° (рис. 189). На гипотенузной поверхности AB напылены узкие равноотстоящие полоски серебра, параллельные преломляющему ребру призмы. Свет падает нормально на одну из боковых граней призмы. Попадая на гипотенузную грань AB, он испытает либо полное отражение от непосеребренных полосок, либо отразится от посеребренных полосок. В последнем случае отражение также практически полное благодаря высокому коэффициенту отражения серебра. В обоих случаях отражение сопровождается изменением фазы, но эти изменения разные (см. § 66 и 73). Таким образом, амплитуда волны при отражении не меняется, а в фазу вносятся периодические изменения.

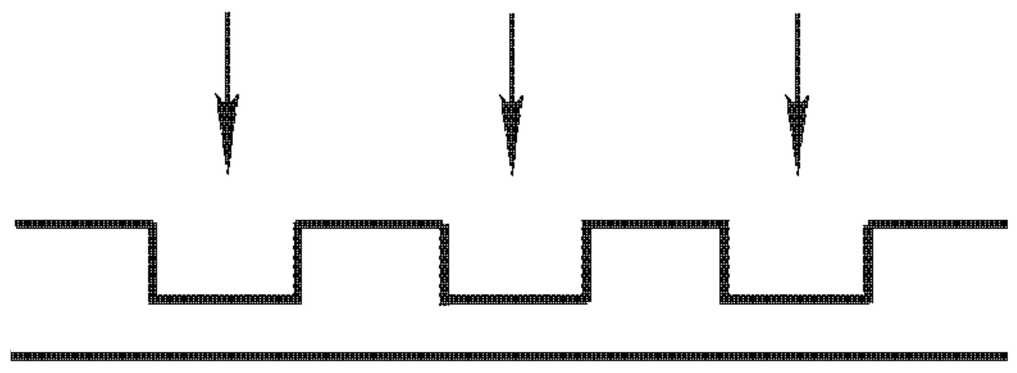


Рис. 188

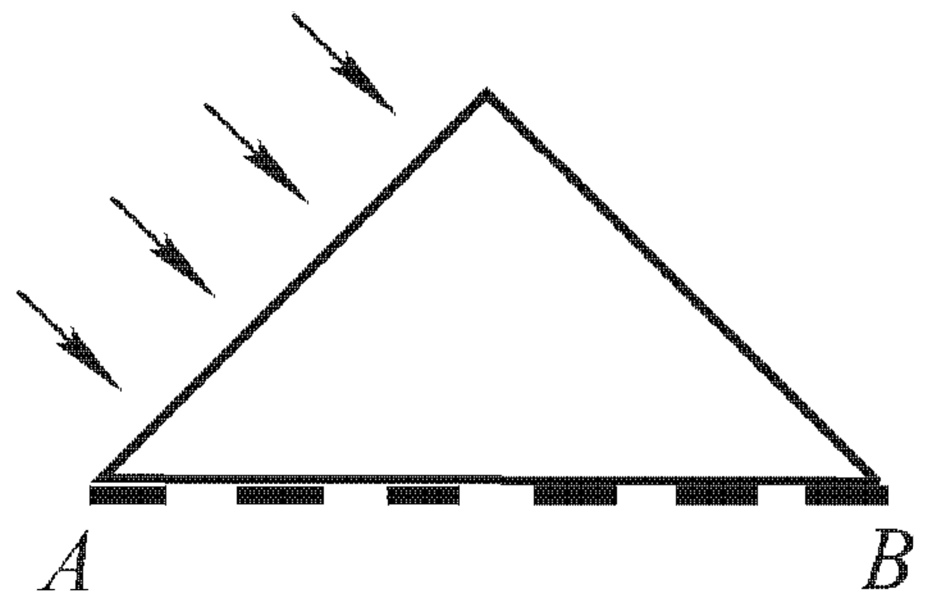


Рис. 189

7. Не все целые числа m, входящие в формулы (46.4) и (46.8), допустимы. Так как $|\sin\vartheta| < 1$, то на m должно быть наложено ограничение

$$\left| m\,\frac{\lambda}{d} + \sin\vartheta_0 \right| \leqslant 1. \tag{46.10}$$

В частности, при нормальном падении

$$|m| \leqslant d/\lambda. \tag{46.11}$$

При $d < \lambda$ может возникнуть только главный максимум нулевого порядка. Плоская волна после прохождения через решетку продолжает оставаться плоской, и никаких боковых волн не возникает. Решетка ведет себя подобно плоскопараллельной пластинке. Но это справедливо только вдали от решетки. В тонком поверхностном слое вблизи решетки существуют неоднородные волны, быстро затухающие при удалении от нее (см. § 52).

8. Допустим, что решетка грубая ($d \gg \lambda$), а свет падает на нее под небольшим углом ϑ_0. Тогда углы дифракции будут малы, так что вместо (46.8) можно написать приближенно

$$d \cdot (\vartheta - \vartheta_0) = m\lambda. \tag{46.12}$$

На очень грубых решетках обнаружить дифракцию трудно из-за малости углов дифракции $(\vartheta - \vartheta_0)$. Но этого можно добиться, применяя скользящее падение света, когда угол падения близок к 90°. При малой разности $(\vartheta - \vartheta_0)$ формулу (46.8) можно переписать в виде

$$d\cos\vartheta_0 \cdot (\vartheta - \vartheta_0) = m\lambda. \tag{46.13}$$

По виду формула (46.13) совпадает с формулой (46.12). Роль периода d играет величина $d\cos\vartheta_0$, которая может быть сделана очень малой. Следовательно, скользящее падение как бы уменьшает период решетки и увеличивает углы дифракции. Таким путем удается получать отчетливые дифракционные спектры даже от очень грубых решеток, например от граммофонных пластинок. Последние позволяют в демонстрационной аудитории получать в белом свете даже довольно красивые дифракционные спектры разных порядков.

Большое значение метод скользящего падения имеет в рентгеновской спектроскопии. Изготовление дифракционной решетки для рентгеновской области спектра встречает большие трудности из-за исключительной малости длин рентгеновских волн (порядка 0,1 нм и меньше). Но метод скользящего падения позволяет получать великолепные дифракционные картины в рентгеновском свете от обычных отражательных оптических решеток. Задача облегчается тем, что показатель преломления рентгеновских лучей меньше единицы. Это позволяет применять такие углы падения, при которых рентгеновские лучи испытывают полное отражение. Таким образом, с помощью оптической дифракционной решетки можно точно измерить длину волны монохроматического рентгеновского излучения. Изучая далее дифракцию этого излучения на естественном кристалле, можно измерить в абсолютных единицах постоянную решетки этого кристалла. После этого такой кристалл может быть использован в рентгеновском спектрографе для измерения длин волн рентгеновских лучей в абсолютных единицах (см. § 61).

ЗАДАЧИ

1. Показать, что для обычной щелевой дифракционной решетки $I_{\text{диф}} \leqslant \leqslant {}^1\!/_4 I_{\text{пад}}$, где $I_{\text{пад}}$ — интенсивность падающего, а $I_{\text{диф}}$ — дифрагированного света, приходящегося на все дифракционные максимумы, кроме максимума нулевого порядка. Максимальное значение интенсивности дифрагированного света достигается, когда ширина щелей решетки равна половине ее периода.

Решение. Если I_0 — интенсивность света, сосредоточенного в максимуме нулевого порядка, а $I_{\text{диф}}$ — дифрагированного света, то $I_{\text{прош}} = I_0 + I_{\text{диф}}$, где $I_{\text{прош}}$ — интенсивность прошедшего света. Аналогично, для дополнительной решетки $I'_{\text{прош}} = I'_0 + I'_{\text{диф}}$. По теореме Бабине $I_{\text{диф}} = I'_{\text{диф}}$. Кроме того,

$$\frac{I_{\text{прош}}}{I'_{\text{прош}}} = \frac{b}{d-b}, \quad \frac{I_0}{I'_0} = \left(\frac{b}{d-b}\right)^2.$$

С учетом этого получаем

$$I_{\text{диф}} = \frac{b}{d}\frac{d-b}{d} I_{\text{пад}}.$$

Максимум этого выражения достигается при $b = {}^1\!/_2 d$, а потому $I_{\text{диф}} \leqslant {}^1\!/_4 I_{\text{пад}}$.

2. При каком условии можно наблюдать зеркальное отражение от шероховатой поверхности при малых и больших углах падения?

Решение. Обозначим через ϑ угол, который падающие лучи образуют с нормалью к плоскости, изображенной на рис. 190 штриховой линией.

Разность хода вторичных волн, исходящих от поверхности тела под углом ϑ' к нормали, определяется выражением

$$\Delta = AD - BC = a(\sin\vartheta' - \sin\vartheta) + h(\cos\vartheta' + \cos\vartheta).$$

Здесь a может принимать любые значения. Поэтому, если $\vartheta' \neq \vartheta$, то Δ может принимать также любые значения, и притом для неправильно отражающей поверхности они с одинаковой вероятностью могут быть и положительными, и отрицательными. Отсюда следует, что правильное отражение, если таковое возможно, может происходить лишь под углом $\vartheta' = \vartheta$. В таком случае $\Delta = 2h\cos\vartheta$. Следовательно, каково бы ни было h, можно подобрать достаточно большой угол ϑ, для которого $\Delta \ll \lambda$. При этом условии отражение будет правильное. При нормальном падении $\Delta = 2h$, и правильное отражение возможно лишь при соблюдении условия $h \ll \lambda$. Таким образом, если угол падения близок к $\pi/2$, зеркальное отражение наблюдается всегда. При малых углах падения зеркальное отражение может наблюдаться, если шероховатость поверхности $\ll\lambda$.

Рис. 190

3. Прозрачная одномерная периодическая структура, профиль которой изображен на рис. 191, освещается сверху плоской монохроматической волной, падающей нормально на верхнюю границу. Ширины уступов и впадин структуры одинаковы. При заданном показателе преломления n подобрать глубину h таким образом, чтобы главные фраунгоферовы дифракционные максимумы первого порядка имели наибольшую интенсивность. Какова при этом интенсивность нулевого главного максимума?

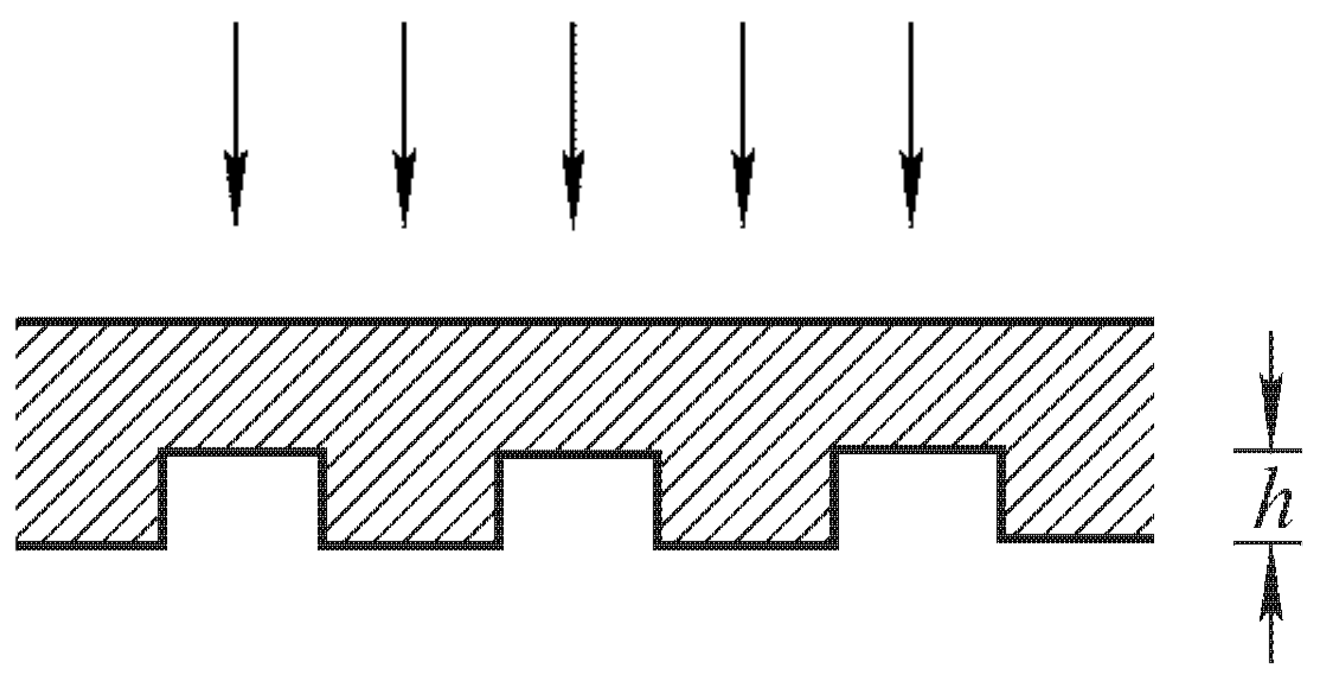


Рис. 191

О т в е т. $h = \dfrac{2m-1}{2(n-1)}\lambda$, где $m = 1, 2, 3, \ldots$

Интенсивность нулевого главного максимума равна нулю.

4. Сложная излучающая система состоит из конечного ряда густо расставленных параллельных вибраторов с равномерным распределением фаз колебаний вдоль ряда. Как должен изменяться со временем сдвиг фаз $\Delta\varphi$ между двумя соседними вибраторами, чтобы главный лепесток диаграммы направленности всей системы совершал круговой обзор местности с постоянной угловой скоростью Ω (при отсутствии вращения самой системы)?

Ответ. Если угол между нормалью к ряду вибраторов и направлением на главный лепесток $\vartheta = \Omega t + \delta$, то

$$\Delta\varphi = -\frac{2\pi}{\lambda}\sin(\Omega t + \delta) + 2m\pi,$$

где

$$m = 0, \pm 1, \pm 2, \ldots$$

5. Над центром граммофонной пластинки O_1 помещен точечный источник монохроматического света на высоте $h_1 = 1$ см. Глаз наблюдателя расположен на высоте $h_2 = 10$ см на расстоянии $a = 110$ см от оси пластинки. Помимо геометрического изображения источника глаз видит систему дифракционных полос на поверхности пластинки. Определить расстояние Δx между ними, если расстояние между бороздками $d = 0{,}5$ мм. Длина световой волны $\lambda = 550$ нм.

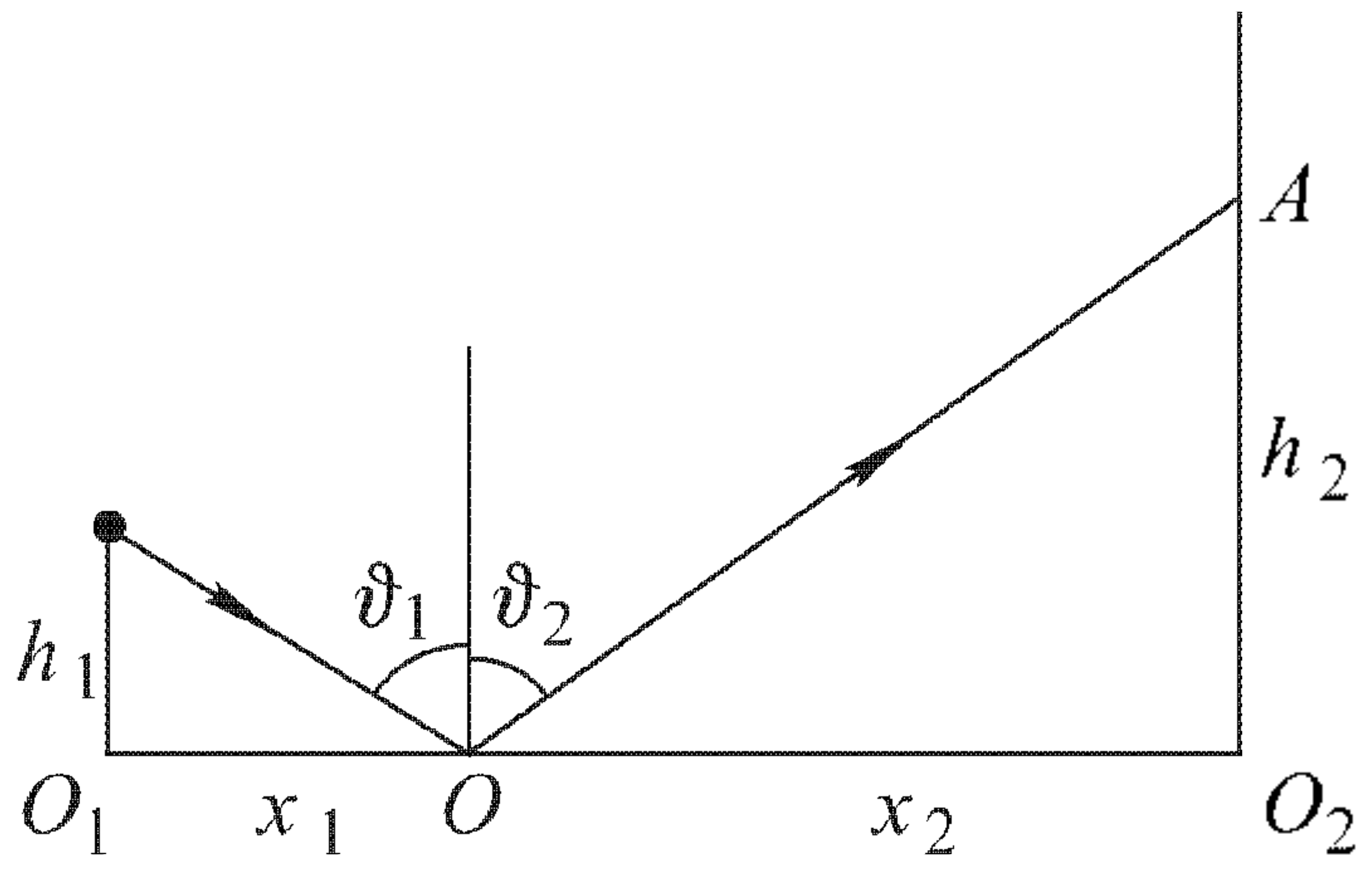


Рис. 192

Решение. Условие максимума m-го порядка (рис. 192) $d(\sin\vartheta_1 - \sin\vartheta_2) = m\lambda$, а максимума $(m+1)$-го порядка $d(\sin\vartheta_1' - \sin\vartheta_2') = (m+1)\lambda$. При переходе от одного максимума к другому углы ϑ_1 и ϑ_2 получают приращения $\Delta\vartheta_1$ и $\Delta\vartheta_2$, связанные соотношением $d(\cos\vartheta_1\,\Delta\vartheta_1 - \cos\vartheta_2\,\Delta\vartheta_2) = \lambda$. Кроме того, $x_1 = h_1 \operatorname{tg}\vartheta_1$, $x_2 = h_2 \operatorname{tg}\vartheta_2$, причем $x_1 + x_2 = a = \text{const}$. Из этого условия получается второе соотношение

$$\frac{h_1}{\cos^2\vartheta_1}\Delta\vartheta_1 + \frac{h_2}{\cos^2\vartheta_2}\Delta\vartheta_2 = 0.$$

Из этих двух соотношений находим $\Delta\vartheta_1$ и $\Delta\vartheta_2$, а после этого — расстояние между максимумами:

$$\Delta x = \Delta x_1 = \Delta x_2 = \frac{h_1}{\cos^2\vartheta_1}\Delta\vartheta_1 = -\frac{h_2}{\cos^2\vartheta_2}\Delta\vartheta_2.$$

Результат можно упростить, заметив, что углы ϑ_1 и ϑ_2 мало отличаются от угла падения ϑ_0, соответствующего зеркальному отражению. Заменив эти углы на ϑ_0, находим

$$\Delta x = \frac{h_1 h_2}{h_1 + h_2}\,\frac{\lambda}{d\cos^3\vartheta_0}.$$

При этом

$$\cos\vartheta_0 \approx \frac{h_1}{x_1} \approx \frac{h_2}{x_2}, \quad \text{или} \quad \cos\vartheta_0 = \frac{h_1 + h_2}{x_1 + x_2} = \frac{h_1 + h_2}{a} = \frac{1}{10}.$$

Подстановка числовых значений дает $\Delta x = 1$ см.

§ 47. Дифракционная решетка как спектральный прибор

1. Положение главных максимумов в дифракционной решетке зависит от длины волны. Исключение составляют только главные максимумы нулевого порядка ($m = 0$), положения которых от длины волны не зависят. Белый и всякий сложный свет можно рассматривать как суперпозицию монохроматических волн с различными длинами. Эти волны при дифракции на решетке ведут себя *независимо.* Поэтому решетка в каждом порядке $m \neq 0$ разложит падающий свет в спектр, в котором отдельные монохроматические компоненты окажутся пространственно разделенными. Главные дифракционные максимумы, соответствующие $m = 1$, образуют *спектр первого порядка.* За ним идет *спектр второго* ($m = 2$), *третьего* ($m = 3$) и *высших порядков.* Если падающий свет белый, то спектр каждого порядка имеет вид цветной полосы, в которой встречаются все цвета радуги. В такой полосе наиболее отклоненными будут красные лучи, наименее отклоненными — фиолетовые.

Положение спектральных линий в спектрах дифракционной решетки определяется простыми соотношениями (46.4) или (46.8). В этом отношении дифракционные спектры выгодно отличаются, например, от спектров призматических, получаемых разложением света дисперсионными призмами. В призматических спектрах положение спектральной линии определяется сложной зависимостью показателя преломления материала призмы от длины волны. Спектр называется *нормальным*, если координата x, характеризующая положение спектральной линии в спектре, линейно меняется с длиной волны. При малых углах дифракции, когда изменением косинуса угла ϑ можно пренебречь, дифракционная решетка дает нормальный спектр.

Важными характеристиками дифракционной решетки и других спектральных аппаратов являются *угловая дисперсия*, *дисперсионная область* и *разрешающая способность.*

2. Угловая дисперсия. *Угловой дисперсией называется производная* $d\vartheta/d\lambda$. Чем больше угловая дисперсия, тем больше расстояние в спектре между двумя спектральными линиями с фиксированными длинами волн. Дифференцируя формулу (46.8) при постоянном ϑ_0, находим для решетки

$$\frac{d\vartheta}{d\lambda} = \frac{m}{d\cos\vartheta} = \frac{\sin\vartheta - \sin\vartheta_0}{\lambda\cos\vartheta}. \tag{47.1}$$

Следовательно, *угловая дисперсия не зависит от параметров решетки, а определяется, помимо длины волны, только углами* ϑ *и* ϑ_0. При заданных d и m она возрастает с увеличением угла ϑ. Наличие косинуса в знаменателе объясняет выгоду скользящего падения (см. конец предыдущего параграфа).

3. Дисперсионная область. Если спектры соседних порядков перекрываются, то спектральный аппарат становится непригодным для исследования соответствующего участка спектра. *Максимальная ширина спектрального интервала* $\Delta\lambda$, *при* *которой еще нет перекрытия, называется дисперсионной областью спектрального аппарата.* Найдем дисперсионную область для дифракционной решетки. Пусть длины волн падающего излучения лежат в спектральном интервале от λ до $\lambda' = \lambda + \Delta\lambda$. Пусть правый конец спектра $(m+1)$-го порядка для длины волны λ совпадает по своему положению с левым концом спектра m-го порядка для длины волны λ' (см. рис. 182). Тогда

$$d(\sin\vartheta - \sin\vartheta_0) = m\lambda',$$
$$d(\sin\vartheta - \sin\vartheta_0) = (m+1)\lambda.$$

Отсюда $m\lambda' = (m+1)\lambda$, а следовательно,

$$\lambda' - \lambda \equiv \Delta\lambda = \lambda/m. \tag{47.2}$$

Это и есть дисперсионная область дифракционной решетки в рассматриваемом участке спектра. При заданной длине волны она определяется только порядком спектра m. Чем больше m, тем у́же дисперсионная область. В дифракционных решетках используются спектры *низких порядков* (обычно второго или третьего). Поэтому дифракционные решетки характеризуются широкими областями дисперсии и пригодны для исследования *широких участков спектра.* В этом основное преимущество дифракционных решеток перед интерференционными спектральными аппаратами, у которых из-за высоких порядков m дисперсионные области очень узкие.

4. Разрешающая способность. Большая дисперсия еще не означает, что две спектральные линии с близкими длинами волн λ и $\lambda' = \lambda + \delta\lambda$ разрешаются спектральным аппаратом, т.е. при их наблюдении воспринимаются как раздельные спектральные линии. Каждая спектральная линия, как бы узка она ни была, изображается спектральным аппаратом не в виде линии, а в виде более или менее размытой дифракционной картины с максимумами и минимумами интенсивности. Дисперсия определяет расстояние, на которое спектральный аппарат разводит центры дифракционных картин, возникающих от двух спектральных линий с различными длинами волн. Если сами картины размыты и имеют значительную ширину, то даже при сравнительно большом разведении их результирующая картина, возникающая от их наложения, неотличима от дифракционной картины, возникающей от одиночной спектральной линии. Чем у́же дифракционные картины от двух близких спектральных линий, тем на меньшее расстояние требуется развести их центры, чтобы разрешить эти спектральные линии. *Наименьшая разность длин волн двух спектральных линий $\delta\lambda$, при которой спектральный аппарат разрешает эти линии, называется спектральным разрешаемым расстоянием, а величина $R = \lambda/\delta\lambda$ — разрешающей способностью аппарата.*

Разрешаемое спектральное расстояние $\delta\lambda$ относится к числу не вполне точно определенных понятий и может быть указано лишь ориентировочно. Для дифракционной решетки Рэлей предложил следующий критерий спектрального разрешения. *Спектральные линии с близкими длинами волн λ и λ' считаются разрешенными, если главный максимум дифракционной картины для одной длины волны совпадает по своему положению с первым дифракционным минимумом в том же порядке для другой длины волны.* Если такой критерий выполняется, то на основании формулы (46.9) можно написать

$$d(\sin\vartheta - \sin\vartheta_0) = \left(m + \frac{1}{N}\right)\lambda,$$

$$d(\sin\vartheta - \sin\vartheta_0) = m\lambda'.$$

Отсюда $(m + 1/N)\lambda = m\lambda'$, и следовательно, $\delta\lambda \equiv \lambda' - \lambda = \lambda/(Nm)$,

$$R = \frac{\lambda}{\delta\lambda} = Nm. \tag{47.3}$$

С целью обоснования критерия Рэлея начертим распределение интенсивности для спектральных линий λ и λ', сдвинутых одна относительно другой на расстояние $\delta\lambda$ (рис. 193). Так как спектральные линии с различными длинами волн некогерентны, то результирующая интенсивность найдется простым сложением интенсивностей обеих линий. Она представлена на рис. 193 сплошной кривой, когда интенсивности обеих спектральных линий одинаковы, а расстояние между ними $\delta\lambda$ соответствует критерию Рэлея. В центре кривой получился провал или минимум, интенсивность в котором, как легко подсчитать, составляет около 80 % максимальной. При наличии такого провала нормальный глаз обычно воспринимает получившуюся картину как «двойную спектральную линию».

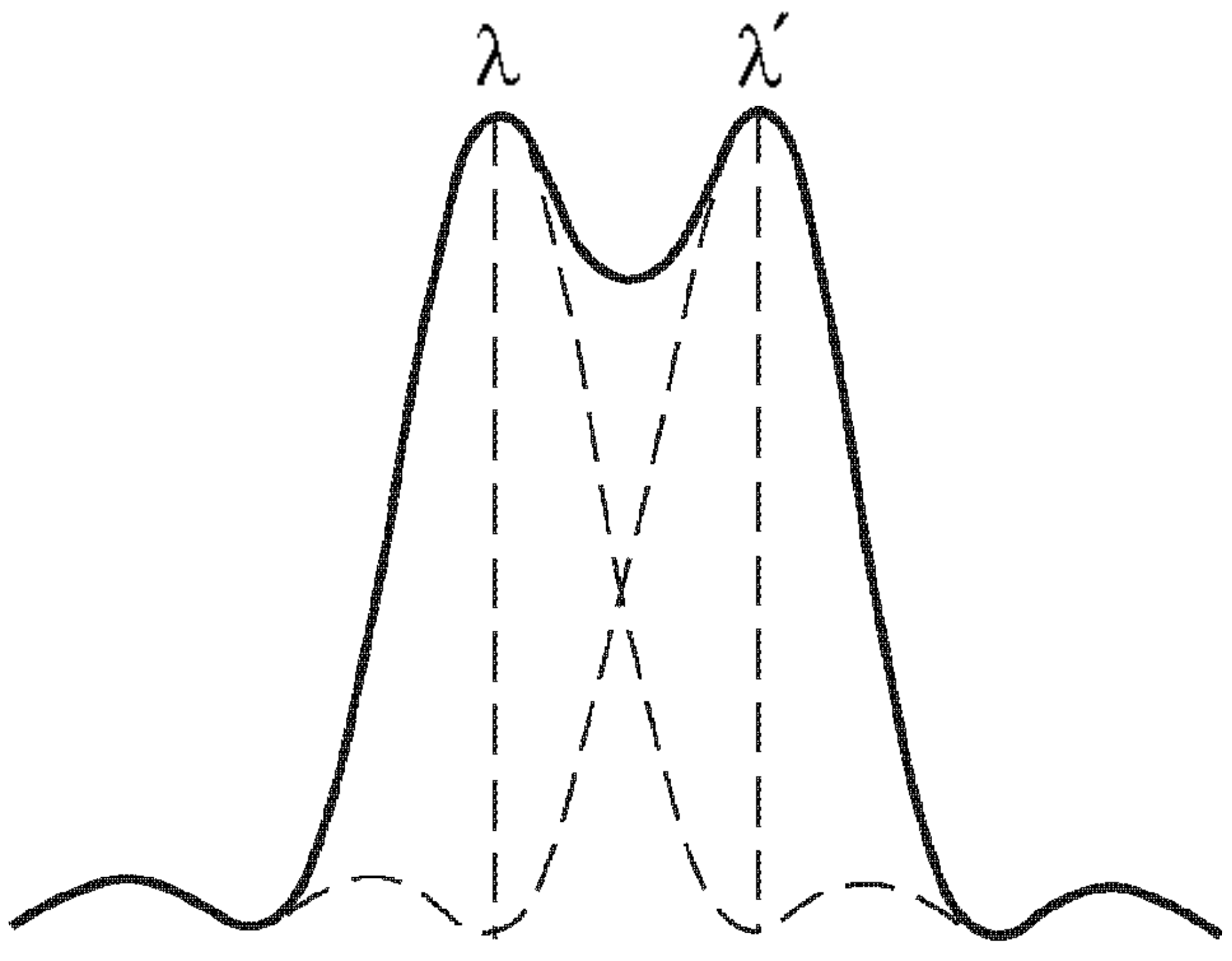


Рис. 193

5. Для повышения разрешающей способности можно либо увеличивать число штрихов N, либо повышать порядок интерференции m. Первый путь используется в дифракционных решетках, второй — в интерференционных спектральных приборах. В Советском Союзе изготовляются плоские и вогнутые дифракционные решетки различных размеров и с различным числом штрихов на миллиметр. Для ультрафиолетовой и видимой областей изготовляются дифракционные решетки, имеющие 1200 и 600 штрихов на миллиметр при размерах

100×100 мм^2 и 150×150 мм^2, а для инфракрасной области — от 300 до 1 штриха на миллиметр при размерах от 150×150 мм^2 до 300×300 мм^2 Таким образом, общее число штрихов доходит приблизительно до 200 000, а разрешающая способность во втором порядке — до 400 000. Важным достоинством дифракционной решетки является малый порядок спектра m. Благодаря этому дифракционные решетки обладают широкими дисперсионными областями $\Delta\lambda = \lambda/m$ и пригодны для исследования *широких интервалов спектра*. Недостатками дифракционных решеток являются малая светосила и сложность в обращении.

В интерференционных спектральных приборах число интерференционных пучков N относительно невелико (несколько десятков, в интерферометре Майкельсона $n = 2$), тогда как порядки спектров m очень высоки (около 10 000 и больше). Поэтому интерференционные спектральные приборы имеют малые дисперсионные области. Они могут применяться для исследования только *очень узких участков спектра*, например для изучения структуры отдельных спектральных линий, выделенных каким-либо другим спектральным аппаратом с большей дисперсионной областью, но с недостаточной разрешающей способностью. Однако эти приборы более просты в обращении и имеют бóльшую светосилу, чем дифракционные решетки.

6. К решетке, как точному спектральному прибору, предъявляются очень высокие требования. Надо нанести десятки или сотни тысяч совершенно идентичных штрихов с идеальной периодичностью. Поэтому техника изготовления дифракционных решеток совершенствовалась довольно медленно. Первая дифракционная решетка, по-видимому, была изготовлена в 1785 г. американским астрономом Риттенгаузом. Но ни самим Риттенгаузом и никем другим она не использовалась для получения и изучения спектров. Решетка была вновь открыта в 1821 г. Фраунгофером, который заложил основы дифракции в параллельных лучах и выполнил первые исследования с помощью дифракционного спектроскопа (в частности, открыл темные линии в спектре солнечного излучения). Фраунгофер построил решетки с числом штрихов от 15 до 130 на сантиметр, наматывая тонкую проволоку на два параллельных винта. До недавнего времени подобные проволочные решетки применялись в области длинных (инфракрасных) волн. Затем Фраунгофер стал изготовлять более совершенные решетки штриховкой слоя золота на поверхности стекла и, наконец, штриховкой самого стекла алмазным острием. Наилучшие решетки были получены последним методом. Лучшая решетка имела ширину около 12 мм и период $3 \cdot 10^{-3}$ мм (3300 штрихов на сантиметр).

После Фраунгофера многие искусные механики уделяли много внимания штриховке решеток. Особо следует отметить астронома-любителя Л.М. Резерфорда, большая часть решеток которого была изготовлена в 1880 г. Его решетки значительно превосходили все предшествующие. Резерфорд ввел в практику отражательные решетки, нанося делительной машиной штрихи на зеркальной поверхности металла. Металл более мягок, чем стекло, и поэтому значительно меньше изнашивает алмазное острие, от постоянства которого так сильно зависит качество решетки.

Но наиболее значительные усовершенствования были сделаны Роулэндом (1848–1901). Он усовершенствовал способы изготовления винтов для делительной машины и первый стал изготовлять вогнутые отражательные решетки, выполняющие одновременно функции решетки и собирающей линзы. Решетки Роулэнда имели до 8000 штрихов на сантиметр при ширине до 10 см и превосходном качестве. Они сделали возможным выполнение важнейших спектроскопических исследований в конце XIX и начале XX веков.

Дальнейшие улучшения в машины Роулэнда ввели Андерсон и Вуд, которые после смерти Роулэнда заменили его в его лаборатории. Из этой лаборатории и поныне выходят наиболее совершенные решетки, имеющие до 12 000 штрихов на сантиметр. С таких гравированных решеток получают дешевые копии (реплики) путем изготовления отпечатков на желатине или специальных пластмассах. В СССР налажено производство дифракционных решеток, а также реплик высокого качества. Их параметры были указаны выше.

§ 48. Эшелон Майкельсона и интерференционные спектральные приборы

1. Дифракционная решетка, как уже указывалось, есть спектральный прибор, в котором осуществляется многолучевая интерференция идентичных световых пучков, из которых каждый сдвинут по фазе относительно предыдущего на одну и ту же величину. Ко всем приборам, работающим по этому принципу, применимы общие результаты, изложенные в двух предыдущих параграфах. В частности, их спектральная разрешающая способность выражается формулой $\lambda/\delta\lambda = Nm$, т.е. равна *разности хода между крайними интерферирующими лучами, выраженной в длинах волн.* Высокая разрешающая способность достигается как увеличением числа интерферирующих пучков N, так и повышением порядка интерференции m. В дифракционных решетках число интерферирующих пучков очень велико (до 200 000 и больше), а порядки интерференции m низкие (не выше 3). В так называемых *интерференционных спектральных приборах*, наоборот, порядки интерференции m очень высокие (до 10^5 и выше), а число интерферирующих пучков сравнительно невелико (20–40). Одним из таких приборов является ступенчатая решетка, или *эшелон Майкельсона.*

2. Эшелон Майкельсона состоит из нескольких (30–40) пластин из очень однородного стекла, толщина h которых порядка 1–3 см. Толщина пластин должна быть совершенно одинакова с точностью до сотой длины волны. Поэтому их вырезают из одной и той же плоскопараллельной пластины, постоянство толщины которой контролируется интерференционными методами. Пластины сложены в виде ступенчатой лестницы (рис. 194).

Ширина ступеньки a должна быть одной и той же на протяжении всей «лестницы». Получается как бы сплошной кусок однородного стекла, ступенчатая поверхность которого действует как дифракционная решетка. Пусть на эшелон нормально падает пучок параллельных лучей. Разность хода между вторичными волнами Гюйгенса, исходящими под углом ϑ от соседних ступенек эшелона и приходящими в соответствующую точку фраунгоферовой дифракционной картины, будет $(BDF) - AE = nh + a\sin\vartheta - h\cos\vartheta$ (рис. 195). Положения главных максимумов определяются условием

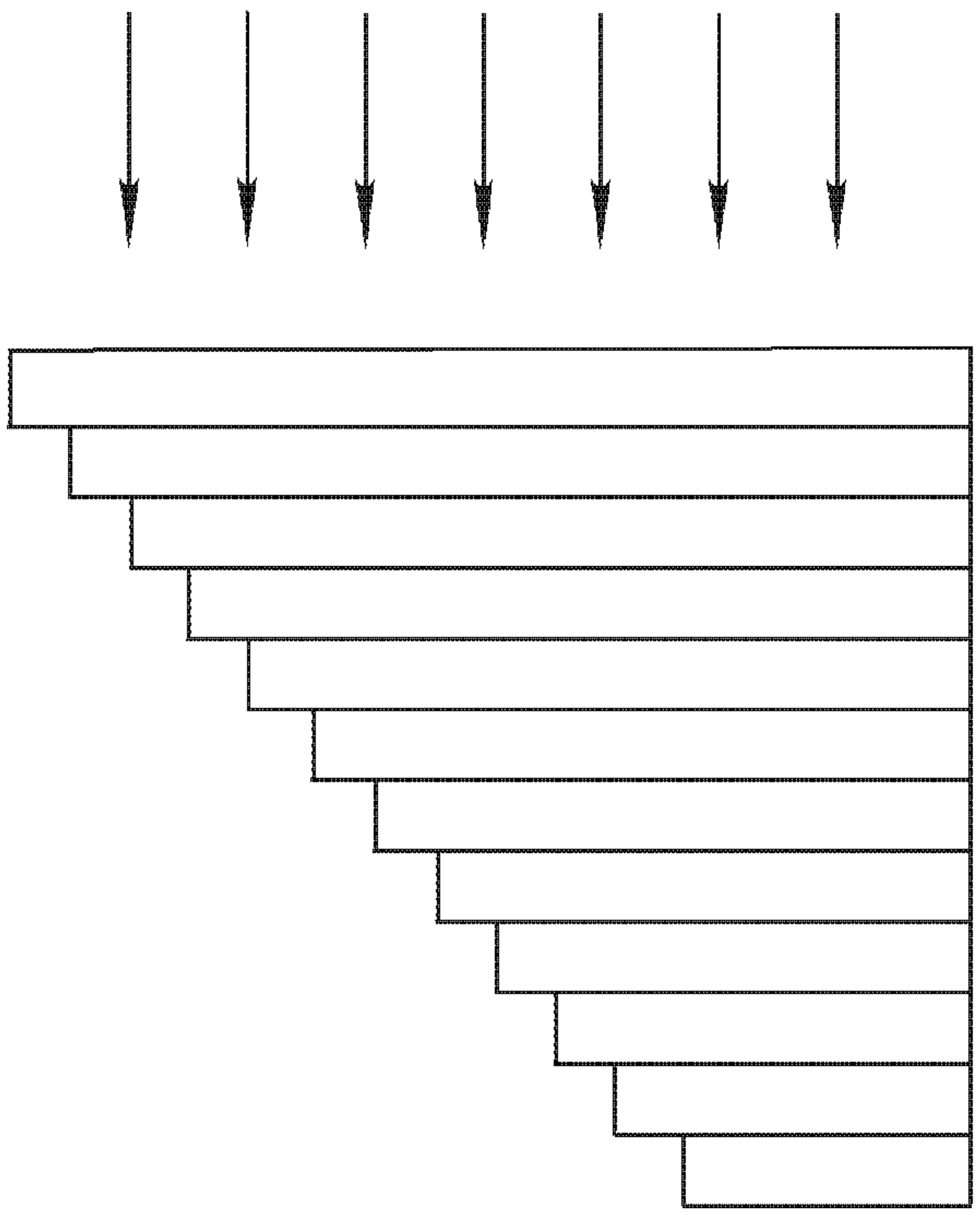


Рис. 194

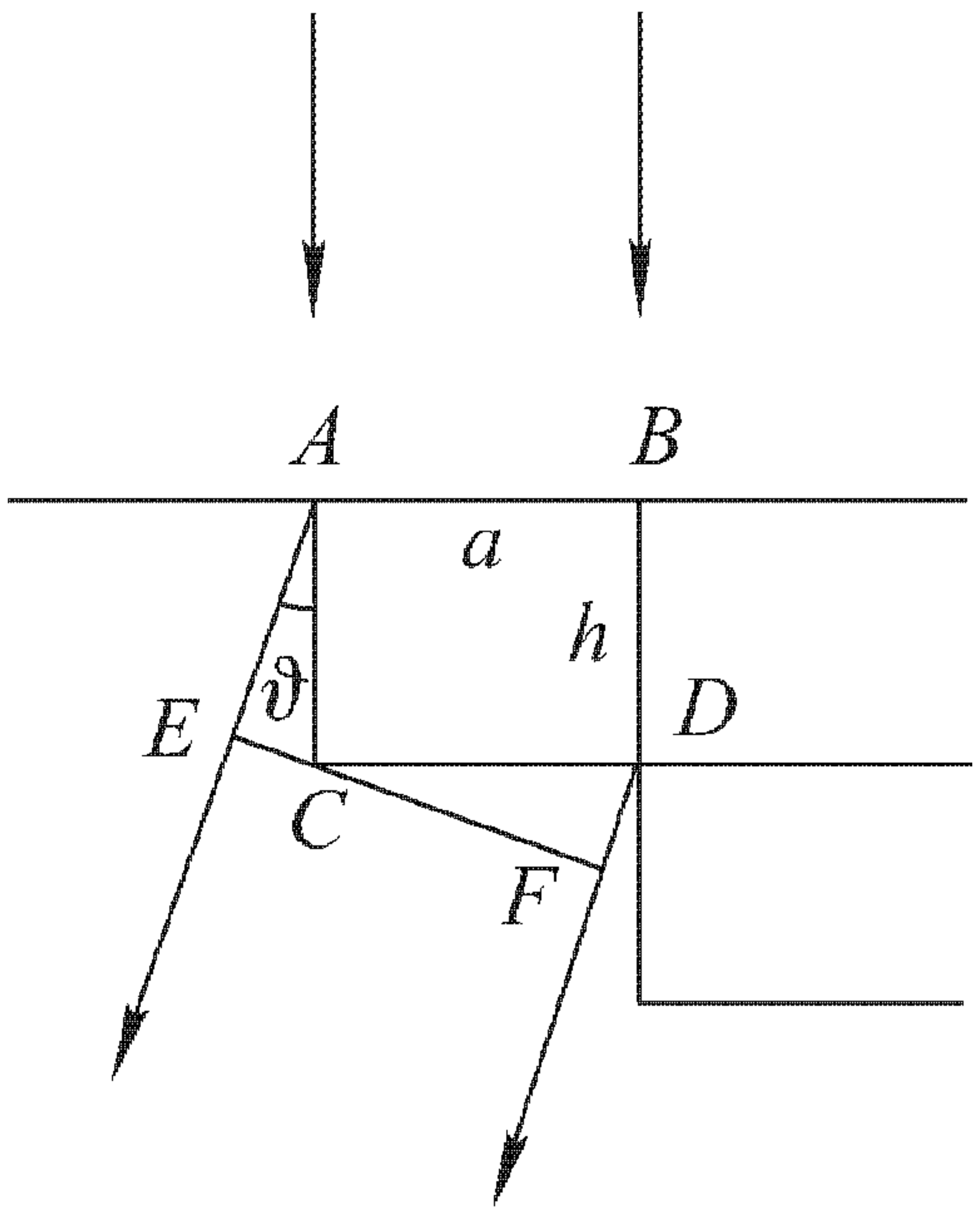


Рис. 195

$$nh + a\sin\vartheta - h\cos\vartheta = m\lambda, \tag{48.1}$$

где m — целое число.

Из (48.1) дифференцированием находим угловую дисперсию

$$\frac{d\vartheta}{d\lambda} = \frac{m}{a\cos\vartheta + h\sin\vartheta}, \tag{48.2}$$

или, ввиду малости угла ϑ, $d\vartheta/d\lambda = m/a$, или

$$\frac{d\vartheta}{d\lambda} = \frac{h(n-1)}{a\lambda}. \tag{48.3}$$

Так как $h \sim a$, то угловая дисперсия эшелона относительно велика. Поэтому при незначительных изменениях λ получаются заметные изменения угла дифракции ϑ.

Дисперсионная область эшелона определяется выражением

$$\Delta\lambda = \frac{\lambda}{m} = \frac{\lambda^2}{h(n-1)}. \tag{48.4}$$

Она очень мала, и это является недостатком эшелона. Например, при $h = 1$ см, $n = 1{,}5$, $\lambda = 600$ нм получаем $\Delta\lambda \approx 0{,}07$ нм, что примерно в 10 раз меньше расстояния между компонентами желтой D-линии натрия.

Разрешающая способность дается формулой

$$\frac{\lambda}{\delta\lambda} = Nm = \frac{Nh(n-1)}{\lambda}, \tag{48.5}$$

где N — число пластин. При $N = 40$ в предыдущем примере получим $\lambda/\delta\lambda \approx 3{,}3 \cdot 10^5$. Со штриховой решеткой в спектре первого порядка такой разрешающей способности можно было бы достигнуть при числе штрихов порядка 300 000.

Уточним формулу (48.5), приняв во внимание дисперсию показателя преломления стекла. Пусть при угле дифракции ϑ получается дифракционный максимум m-го порядка для длины волны $\lambda' = \lambda + \delta\lambda$. Тогда

$$hn' + a\sin\vartheta - h\cos\vartheta = m\lambda' = m(\lambda + \delta\lambda),$$

где n' — показатель преломления стекла для той же длины волны. Согласно Рэлею, $\delta\lambda$ будет равна разрешаемому спектральному расстоянию, если для длины волны λ в том же направлении получится первый дифракционный минимум, т.е.

$$hn + a\sin\vartheta - h\cos\vartheta = m\lambda + \lambda/N.$$

Вычитая это соотношение из предыдущего, получим

$$\frac{\lambda}{\delta\lambda} = N\left(m - h\,\frac{dn}{d\lambda}\right), \tag{48.6}$$

или после подстановки $m = h(n-1)/\lambda$

$$\frac{\lambda}{\delta\lambda} = \frac{Nh}{\lambda}\left[(n-1) - \lambda\,\frac{dn}{d\lambda}\right]. \tag{48.7}$$

Так как в прозрачной области спектра $dn/d\lambda < 0$ (нормальная дисперсия), то дисперсия показателя преломления повышает разрешающую способность эшелона.

Рассуждая аналогично, легко получить следующие выражения для дисперсионной области и угловой дисперсии эшелона:

$$\Delta\lambda = \frac{\lambda}{m - h(dn/d\lambda)}, \tag{48.8}$$

$$\frac{d\vartheta}{d\lambda} = \frac{m - h(dn/d\lambda)}{a\cos\vartheta + h\sin\vartheta}. \tag{48.9}$$

Дисперсия показателя преломления увеличивает угловую дисперсию эшелона, но уменьшает его дисперсионную область.

Интенсивность света в дифракционной картине, разумеется, определяется общей формулой (46.2). В ней $I_1(\vartheta)$ означает интенсивность, которая получилась бы при дифракции на отдельной ступеньке эшелона ширины a. Функция $I_1(\vartheta)$ такая же, что и при дифракции на щели. Поэтому весь дифрагированный свет практически концентрируется в интервале углов ϑ между $\vartheta_1 = -\lambda/a$ и $\vartheta_2 = +\lambda/a$. Заметную интенсивность будут иметь только такие главные максимумы, которые попадают в указанный интервал. Ввиду малости угла ϑ можно пренебречь его квадратами, представив формулу (48.1) в виде $(n-1)h + a\vartheta = m\lambda$, или

$$\vartheta = \frac{m\lambda - (n-1)h}{a}. \tag{48.10}$$

Допустим, что показатель преломления таков, что для определенного m числитель в этой формуле обращается в нуль. Тогда в указанный интервал попадут только три главных максимума, порядки которых равны $(m-1)$, m, $(m+1)$. Крайние из них приходятся на концы интервала, их интенсивность равна нулю. Поэтому дифрагированный свет будет состоять только из спектра одного порядка m. Если же числитель в формуле (48.10) не равен нулю, то будут видны спектры двух соседних порядков, вообще говоря, разной интенсивности. Их интенсивности будут одинаковы, когда числитель в формуле (48.10) по абсолютной величине равен $\lambda/2$. Таким образом, возможны такие установки эшелона, при одной из которых получается только один дифракционный максимум, а при другой — два одинаковых симметричных максимума. Чтобы перейти от одной установки к другой, эшелон обычно помещают в герметическую камеру и изменяют в ней давление, а с ним и относительный показатель преломления n, пока не будут достигнуты требуемые для работы условия.

Майкельсону принадлежит также идея использования эшелона как отражательной решетки. Такой эшелон был построен Вильямсом в 1933 г. из пластин плавленого кварца. Тщательно очищенные пластины кварца сажались на оптический контакт и нагревались до температуры, значительно меньшей температуры плавления кварца. В результате пластины прилипали друг к другу настолько прочно, что после охлаждения их удавалось отделить одна от другой только при значительном усилии. Отражающие ступеньки эшелона покрывались алюминием, напыленным в вакууме.

3. В эшелоне Майкельсона интерферирующие пучки возникают в результате дифракции на ступеньках. В чисто интерференционных спектроскопах — интерферометре Фабри–Перо и пластинке Луммера–Герке — они образуются в результате отражений света. Эти приборы уже были рассмотрены в § 36. Получаемый в них спектр определенного порядка есть не что иное, как интерференционная полоса равного наклона. Дисперсионная область, конечно, определяется общей формулой (47.2). Для интерферометра Фабри–Перо порядок спектра m можно вычислить по формуле $m = 2h/\lambda$, где h — расстояние между отражающими поверхностями интерферометра. Для пластинки

Луммера–Герке то же самое можно сделать по формуле (36.7), заменяя в ней ψ предельным углом полного отражения. Это дает

$$m = \frac{2h\sqrt{n^2-1}}{\lambda}. \tag{48.11}$$

Остановимся еще на разрешающей способности интерферометра Фабри–Перо и пластинки Луммера–Герке. В них интерферируют пучки, интенсивность которых медленно убывает с возрастанием номера пучка. Если бы число пучков было бесконечно, как предполагалось при вычислениях в § 36, то спектр содержал бы только одни главные максимумы (см. рис. 141) и никаких добавочных максимумов и минимумов[1]. В этом случае критерий спектрального разрешения Рэлея теряет смысл. Поэтому в § 36 был дан другой критерий. Для разрешающей способности интерферометра Фабри–Перо он приводит к формуле (36.5), имеющей тот же вид, что и формула (47.3). Роль числа интерферирующих пучков N играет величина $N = 2\pi\sqrt{R}/(1-R)$, практически равная $2\pi/(1-R)$. В этом нет ничего неожиданного, так как интуитивно следует ожидать, что убывание интенсивности эквивалентно ограничению числа эффективно действующих пучков без учета их ослабления. Число таких эффективных пучков, очевидно, пропорционально $1/(1-R)$.

Что касается пластинки Луммера–Герке, то для нее в формулу (36.5) необходимо ввести поправку на зависимость показателя преломления стекла от длины волны λ, а также на конечное число интерферирующих пучков. Поэтому мы повторим вывод формулы (36.5) с этим уточнением, считая сначала число интерферирующих пучков бесконечным. Сохраняя обозначения прежнего вывода, пишем, как и раньше, $\delta\Phi = (1-R)/\sqrt{R}$, причем в максимуме m-го порядка $\Phi = 2\pi m$, а во всех остальных точках $\Phi = (4\pi dn\cos\psi)/\lambda$. Однако теперь величины n и ψ имеют разные значения для λ и λ'. Общим для них является только угол падения (так как свет исследуется вне пластинки), а потому $n\sin\psi = n'\sin\psi' = \sin\varphi$. Следовательно, должно быть

$$\frac{\delta\Phi}{\Phi} = \frac{\delta n}{n} - \frac{\sin\psi\,\delta\psi}{\cos\psi} - \frac{\delta\lambda}{\lambda}$$

при условии

$$\frac{\delta n}{n} + \frac{\cos\psi}{\sin\psi}\,\delta\psi = 0.$$

Исключив $\delta\psi$, получим

$$\frac{\delta\Phi}{\Phi} = \frac{1}{\cos^2\psi}\frac{\delta n}{n} - \frac{\delta\lambda}{\lambda}.$$

---

[1] Реально они, конечно, есть из-за ограниченности числа интерферирующих пучков.

Подставим сюда значения $\delta\Phi$, $\Phi = 2m\pi$, а также $\delta n = (dn/d\lambda)\delta\lambda$. Тогда

$$\left|\frac{\lambda}{\delta\lambda}\right| = \frac{2\pi\sqrt{R}m}{1-R}\left(1 - \frac{1}{\cos^2\psi}\frac{\lambda}{n}\frac{dn}{d\lambda}\right). \tag{48.12}$$

Эта формула выведена в предположении бесконечного числа интерферирующих пучков. Она была бы применима к интерферометру Фабри–Перо, если бы в нем пространство между отражающими зеркалами было заполнено однородной средой, показатель преломления n которой отличен от единицы (для такого интерферометра можно было бы положить $\cos\varphi = 1$). Для пластинки Луммера–Герке формула неприменима из-за ограниченного числа интерферирующих пучков. В этом случае, как ясно из изложенного выше, формулу надо заменить на

$$\left|\frac{\lambda}{\delta\lambda}\right| = Nm\left(1 - \frac{1}{\cos^2\psi}\frac{\lambda}{n}\frac{dn}{d\lambda}\right). \tag{48.13}$$

После подстановки значений N, m и $\cos\psi$ получаем

$$\frac{\lambda}{\delta\lambda} = L\left(\frac{n^2-1}{\lambda} - n\frac{dn}{d\lambda}\right). \tag{48.14}$$

Аналогично, для дисперсионной области

$$\Delta\lambda = \frac{\lambda^2}{2h\sqrt{n^2-1}\left(1 - \dfrac{\lambda n}{n^2-1}\dfrac{dn}{d\lambda}\right)}. \tag{48.15}$$

Из-за дисперсии показателя преломления стекла разрешающая способность пластинки Луммера–Герке возрастает, а дисперсионная область уменьшается.

§ 49. Разрешающая способность призмы

1. Действие призмы как спектрального прибора основано на зависимости показателя преломления вещества от длины волны. Для определения разрешающей способности призмы необходимо учесть дифракцию света на краях диафрагмы или самой призмы, ограничивающих ширину падающего светового пучка. Допустим сначала, что на призму падает монохроматический параллельный пучок лучей, ограниченный диафрагмой AA' (рис. 196). Пусть волновой фронт падающей волны совпадает с плоскостью диафрагмы AA'. Возьмем за призмой произвольный волновой фронт BB'. По определению волнового фронта оптические длины $(ACDB)$ и $(A'C'D'B')$ одинаковы:

$$AC + na + DB = A'C' + nb + D'B', \tag{49.1}$$

где a и b — геометрические длины CD и $C'D'$, проходимые светом в веществе призмы.

Если крайние точки B и B' сместить бесконечно мало вдоль плоскости BB' и соединить их с A и A' бесконечно близкими виртуальными лучами, то в силу принципа Ферма равенство (49.1) сохранится с точностью до членов высших порядков относительно этих боковых смещений. Это приводит к следующему правилу построения волнового фронта. Надо от точек исходного волнового фронта провести действительные или бесконечно близкие к ним виртуальные лучи и отложить на них отрезки одинаковой оптической длины. Геометрическое место концов таких отрезков и даст новое положение волнового фронта. В случае плоских волновых фронтов достаточно ограничиться построением крайних лучей $ACDB$ и $A'C'D'B'$. При построении волновых фронтов, отличающихся бесконечно мало своими направлениями, можно пользоваться одними и теми же световыми путями, хотя истинные пути света и отличаются друг от друга. То же справедливо для световых пучков, бесконечно мало отличающихся длинами волн. Это используется ниже для упрощения вычислений.

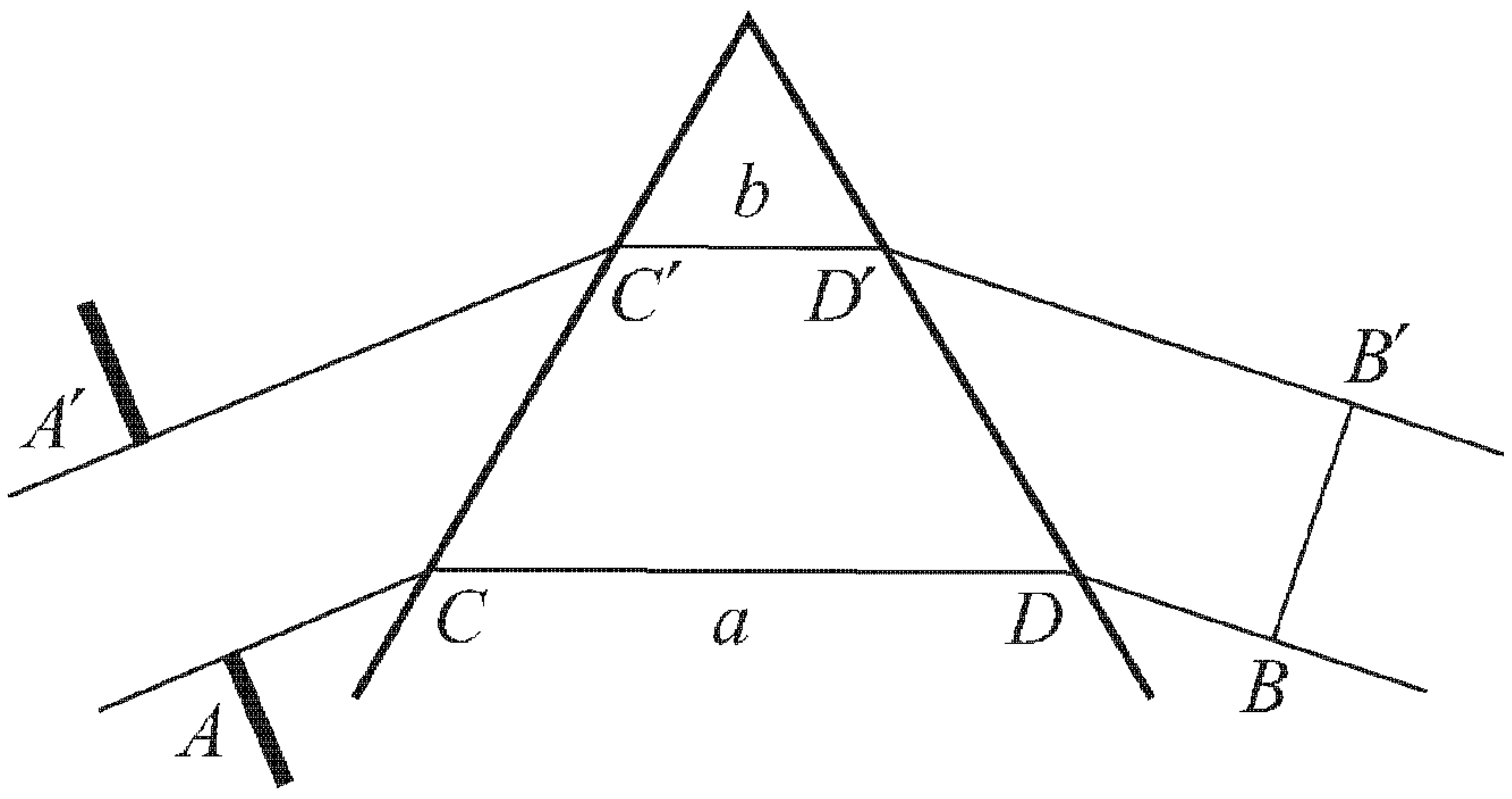


Рис. 196

Учтем теперь дифракцию света на краях диафрагмы AA' и определим за призмой направления на нулевой дифракционный максимум и дифракционные минимумы. Проведем крайние лучи $ACDB$ и $A'C'D'B'$, которые могут быть либо действительными, либо бесконечно близкими к ним виртуальными, и пересечем их произвольной плоскостью BB'. (Мы не предполагаем теперь, что плоскость BB' перпендикулярна к световым лучам.) Если разность оптических длин $(ACDB)$ и $(A'C'D'B')$ равна нулю, то плоскость BB' будет одним из волновых фронтов. Нормаль к ней укажет направление на главный максимум, т.е. максимум нулевого порядка. Если же эти оптические длины отличаются на целое число длин волн, то, как и при дифракции на щели, нормаль к плоскости BB' укажет направление на дифракционный минимум соответствующего порядка. В частности, если оптические длины отличаются на λ, то получится дифракционный минимум первого порядка.

2. Теперь нетрудно вывести формулу для разрешающей способности призмы. Предположим, что на диафрагму падает параллельный пучок света с двумя длинами волн λ и λ'. Пусть плоскость BB' есть волновой фронт для длины волны λ'. Нормаль к ней определяет направление на нулевой максимум для длины волны λ'. Пусть то же направление есть направление на дифракционный минимум первого порядка для длины волны λ. Согласно критерию Рэлея разность $\delta\lambda = \lambda' - \lambda$ есть минимальное расстояние, разрешаемое призмой. На основании сказанного выше

$$AC + n(\lambda')a + DB = A'C' + n(\lambda')b + D'B',$$
$$AC + n(\lambda)a + DB = A'C' + n(\lambda)b + D'B' + \lambda.$$

Вычитая почленно, получим

$$(a-b)[n(\lambda') - n(\lambda)] = -\lambda,$$

или

$$(a-b)\frac{dn}{d\lambda}\delta\lambda = -\lambda.$$

Отсюда для разрешающей способности призмы получаем

$$\frac{\lambda}{\delta\lambda} = -(a-b)\frac{dn}{d\lambda}. \qquad (49.2)$$

Из вывода ясно, что под λ следует понимать длину волны в вакууме. Предполагается, что дисперсия стекла нормальная, т.е. $dn/d\lambda < 0$.

Разрешающая способность призмы зависит только от дисперсии показателя преломления $dn/d\lambda$ и разности путей a и b, проходимых в призме крайними лучами пучка. Поэтому для полного использования разрешающей способности необходимо, чтобы световой пучок покрывал всю боковую поверхность призмы, на которую он падает. В этом случае $b = 0$, а a означает длину основания призмы. Например, для стеклянной призмы из тяжелого флинта в желтой области спектра $dn/d\lambda \approx -956$ см$^{-1}$. Разрешающая способность призмы при $a = 1$ см в указанной области спектра будет $\lambda/\delta\lambda \approx 956$. Это минимальная разрешающая способность, при которой может быть разрешена двойная D-линия натрия ($\lambda = 589{,}0$ нм, $\lambda' = 589{,}6$ нм).

Для сложной призмы, состоящей из нескольких простых призм, поставленных одна за другой, формула (49.2) должна быть заменена на

$$\frac{\lambda}{\delta\lambda} = -\sum_i (a_i - b_i)\frac{dn_i}{d\lambda}. \qquad (49.3)$$

Здесь суммирование производится по всем призмам, составляющим систему.

3. Призму можно рассматривать как предельный случай эшелона Майкельсона (см. рисунки 194 и 195). Допустим, что ширина и высота ступеньки эшелона стремятся к нулю, а общее число ступенек — к бесконечности. При этом последовательно будут исчезать спектры высших порядков, пока не останется только спектр нулевого порядка. Начиная с этого момента, эшелон будет действовать как призма. При $m = 0$ формула (48.6) перейдет в формулу (49.2), так как Nh есть длина основания призмы. Для дисперсионной области в этом случае формула (48.8) дает $\Delta\lambda = \infty$. Этого и следовало ожидать, так как весь свет, выходящий из призмы, концентрируется в спектре нулевого порядка, а потому вопрос о наложении спектров разных порядков не возникает.

ЗАДАЧИ

1. Какую разрешающую силу должен иметь спектральный аппарат для разрешения дублета D-линии натрия ($\lambda_1 = 589{,}0$ нм, $\lambda_2 = 589{,}6$ нм)? Подсчитать минимальное число штрихов решетки, которая может разрешить его в спектре первого порядка. При каком наименьшем основании a можно сделать то же самое с помощью призмы, изготовленной из стекла с дисперсией $dn/d\lambda = -956$ см$^{-1}$?

Ответ. 1) $\lambda/\delta\lambda > 1000$; 2) $N \sim 1000$; 3) $a \approx 1$ см.

2. Изменяется ли разрешающая сила решетки при изменении наклона первичного пучка, падающего на нее?

Ответ. Нет.

3. Изменится ли разрешающая способность и дисперсионная область дифракционной решетки, если, закрепив неподвижно трубу, в которую наблюдаются дифракционные спектры, закрыть через одну щели решетки?

Ответ. Разрешающая способность не изменится. Дисперсионная область уменьшится вдвое.

4. На плоскую отражательную решетку, содержащую $N = 50\,000$ штрихов, нормально падает свет от двойной линии натрия ($\lambda_1 = 589{,}0$ нм, $\lambda_2 = 589{,}6$ нм). Число штрихов, приходящееся на 1 см длины решетки, равно $n = 5000$. Какой максимальный порядок спектра m можно получить от такой решетки и каково минимальное расстояние $\delta\lambda$ между спектральными линиями, которое способна разрешить решетка в указанной области спектра? Спектр максимального порядка фотографируется на фотопластинке с помощью объектива с фокусным расстоянием $f = 50$ см. Какое расстояние Δx между спектральными линиями λ_1 и λ_2 получится на фотопластинке?

Ответ. $m = 3$; $\delta\lambda = \dfrac{\lambda}{Nm} = 0{,}004$ нм; $\Delta x = \dfrac{fm\,\delta\lambda}{\sqrt{d^2 - m^2\lambda^2}} = 1$ мм.

5. Как изменится разрешающая способность и дисперсионная область пластинки Луммера–Герке, если 1) ее толщину увеличить вдвое, а длину оставить без изменения; 2) длину увеличить вдвое, а толщину оставить без изменения?

Ответ. 1) Разрешающая способность не изменится, дисперсионная область уменьшится вдвое. 2) Разрешающая способность увеличится вдвое, дисперсионная область не изменится.

6. Какую минимальную длину должна иметь пластинка Луммера–Герке ($n = 1{,}5$), чтобы разрешить дублетную структуру водородной линии H_α ($\lambda = 656{,}3$ нм)? Разность длин волн линий дублета равна 0,014 нм. Пренебречь величиной $dn/d\lambda$.

Ответ. $L = \dfrac{\lambda^2}{(n^2 - 1)\,\delta\lambda} = 2{,}5$ см.

7. Какая интерференционная картина будет наблюдаться при освещении монохроматическим светом двух скрещенных пластинок Луммера–Герке, т.е. двух пластинок, расположенных одна за другой вдоль прямой линии, но повернутых друг относительно друга на 90°?

Ответ. Будет наблюдаться система пятен, расположенных в узлах прямоугольной сетки.

8. Какой должна быть длина a основания стеклянной призмы, чтобы она имела такую же разрешающую способность, как и пластинка Луммера–Герке длиной $L = 20$ см? Показатель преломления пластинки $n = 1{,}5$; дисперсия показателя преломления призмы $dn_{\text{пр}}/d\lambda = -956$ см$^{-1}$; $\lambda = 600$ нм.

Ответ. $a = -\dfrac{L(n-1)}{\lambda\, dn_{\text{пр}}/d\lambda} = 44$ м.

9. Эшелон Майкельсона состоит из $N = 30$ стеклянных пластинок с показателем преломления $n = 1{,}5$; толщина каждой из них $h = 1$ см. Какова должна быть длина a основания стеклянной призмы, чтобы она имела такую же разрешающую способность, что и рассматриваемый эшелон? Дисперсия показателя преломления призмы $dn_{\text{пр}}/d\lambda = 956$ см$^{-1}$; $\lambda = 600$ нм.

Ответ. $a = \dfrac{Nh(n-1)}{\lambda\, dn_{\text{пр}}/d\lambda} = 2{,}6$ м.

10. Стеклянная призма с основанием $a = 10$ см изготовлена из тяжелого флинта, дисперсия которого в окрестности $\lambda = 600$ нм равна $dn/d\lambda = 1000$ см$^{-1}$. Какую максимальную разрешающую способность может иметь дифракционная решетка, ширина заштрихованной части которой равна длине основания этой призмы? Сравнить разрешающую способность такой решетки с разрешающей способностью призмы.

Решение. Из формулы решетки $d(\sin\vartheta - \sin\vartheta_0) = m\lambda$ следует $m\lambda \leqslant 2d$. Отсюда умножением на $N = a/d$ получаем для максимальной разрешающей способности решетки

$$R_{\text{реш}} = Nm = \frac{2a}{\lambda} = 3\cdot 10^5 \quad \text{и} \quad \frac{R_{\text{реш}}}{R_{\text{пр}}} = \frac{2/\lambda}{dn/d\lambda} = 33.$$

Из этого примера видно, что при одних и тех же геометрических размерах разрешающая способность решетки, вообще говоря, много больше, чем призмы.

11. Оценить ширину коллиматорной щели, при которой практически полностью используется теоретическая разрешающая способность призмы.

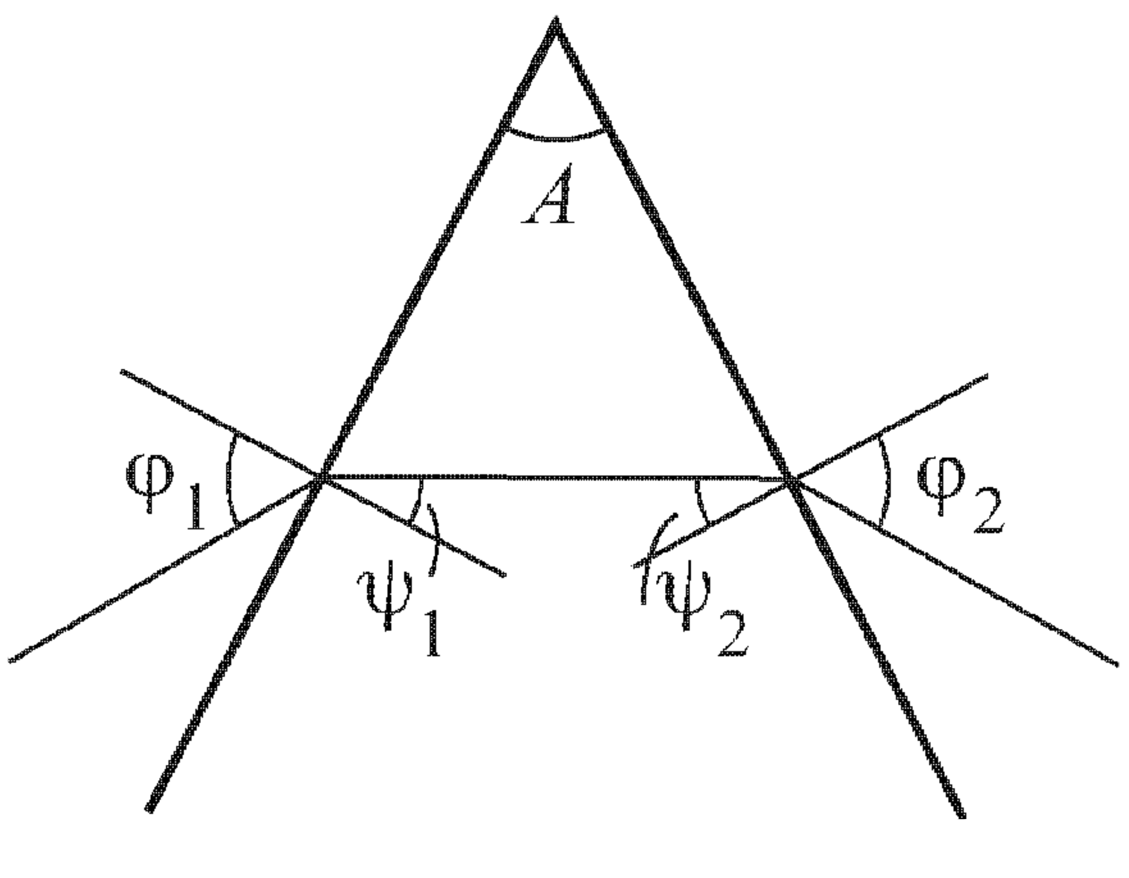


Рис. 197

Решение. Предельная разрешающая способность призмы найдется из формулы (49.2), если положить $b = 0$. Пусть на призму падает плоская волна, содержащая длины волн λ и λ', расстояние между которыми $\delta\lambda$ равно минимальному расстоянию, разрешаемому призмой. Разность показателей преломления для этих длин волн $\delta n = (dn/d\lambda)\delta\lambda$. Подставляя сюда $\delta\lambda = \lambda\Big/\left(a\dfrac{dn}{d\lambda}\right)$, получим $\delta n = \dfrac{\lambda}{a}$. По выходе из призмы первоначально параллельный пучок света из-за дисперсии сделается расходящимся. Рассчитаем угловое расхождение вышедшего пучка. Как видно из рис. 197, $\sin\varphi_1 = n\sin\psi_1$. Дифференцируя при постоянном φ_1, находим: $\delta n\sin\psi_1 + n\cos\psi_1\cdot\delta\psi_1 = 0$. Так как $\psi_1 + \psi_2 = A = \text{const}$ и, следовательно,

$\delta\psi_1 + \delta\psi_2 = 0$, то $\delta n \sin\psi_1 = n\cos\psi_1 \cdot \delta\psi_2$. Из закона преломления $\sin\varphi_2 = n\sin\psi_2$ находим искомое угловое расхождение вышедшего пучка:

$$\delta\varphi_2 = \frac{\sin\psi_2}{\cos\varphi_2}\,\delta n + \frac{n\cos\psi_2}{\cos\varphi_2}\,\delta\psi_2 = \left(\frac{\sin\psi_2}{\cos\varphi_2} + \frac{n\cos\psi_2}{\cos\varphi_2}\,\frac{\sin\psi_1}{n\cos\psi_1}\right)\delta n.$$

При работе призма должна быть установлена на угол наименьшего отклонения, а потому $\varphi_1 = \varphi_2 = \varphi$, $\psi_1 = \psi_2 = \psi$,

$$\delta\varphi_2 = 2\,\frac{\sin\psi}{\cos\varphi}\,\delta n = 2\,\frac{\sin\psi}{\cos\varphi}\,\frac{\lambda}{a}.$$

Если h — ширина щели, f — фокусное расстояние коллиматора, то угловая ширина щели будет $\alpha = h/f$. Для полного использования разрешающей способности призмы угол α должен быть мал по сравнению с углом $\delta\varphi_2$. Практически достаточно, чтобы угол α был меньше приблизительно половины угла $\delta\varphi_2$. Это дает

$$h < \frac{\sin\psi}{\cos\varphi}\,\frac{f\lambda}{a}. \tag{49.4}$$

Так как $\psi = A/2$, то

$$h < \frac{\sin(A/2)}{\sqrt{1 - n^2\sin^2(A/2)}}\,\frac{f\lambda}{a}. \tag{49.5}$$

При $a = 10$ см, $A = 60^\circ$, $n = 1{,}73$, $f = 25$ см, $\lambda = 500$ нм последняя формула дает $h < 1{,}2 \cdot 10^{-3}$ мм.

§ 50. Действие спектрального аппарата на световые импульсы

1. В предыдущих параграфах спектральный аппарат рассматривался как *анализатор*, пространственно разделяющий монохроматические волны, уже имеющиеся (согласно теореме Фурье) в световом возмущении, попадающем в этот прибор. Но на него можно также смотреть как на *генератор*, который под действием непериодических возмущений (импульсов) вырабатывает периодические возмущения и пространственно разделяет их. Проще всего такой подход разъяснить на примере дифракционной решетки.

Предположим, что решетка состоит из бесконечного числа равноотстоящих щелей в непрозрачном экране. Пусть на нее нормально падает какое-либо плоское непериодическое возмущение. Его форма для последующих рассуждений не имеет значения. Для наглядности будем изображать его в виде бесконечно короткого прямоугольного импульса (рис. 198). Достигнув в момент t_0 всех точек решетки, импульс возбудит вторичные цилиндрические волны, исходящие от щелей решетки. Вместо одиночного импульса за решеткой получится бесконечное множество следующих друг за другом пространственно разделенных импульсов с цилиндрическими волновыми фронтами (рис. 199). Сечения волновых фронтов плоскостью чертежа в любой фиксированный момент времени t будут окружностями одного и того же радиуса $c(t - t_0)$, описанными из щелей решетки как из центров (рис. 200). На больших расстояниях от решетки малые участки волновых фронтов могут рассматриваться как плоские и следующие друг за другом через равные промежутки времени

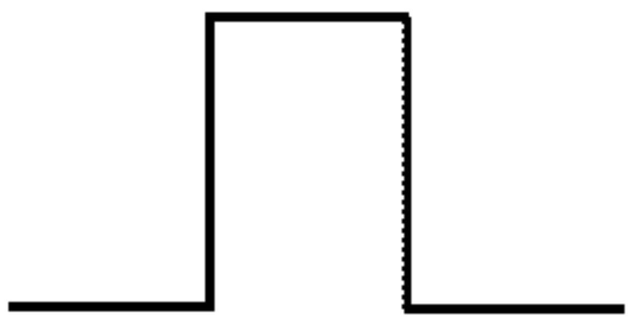


Рис. 198

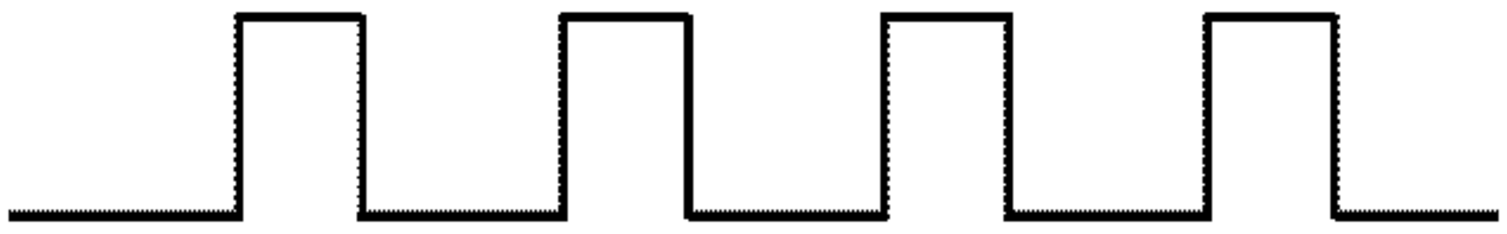


Рис. 199

$$\tau = \frac{d \sin\vartheta}{c}. \tag{50.1}$$

На таких расстояниях возмущение представляется периодической функцией $E(t)$, период которой τ зависит от угла ϑ, т.е. от направления излучения. Вместо непериодического импульса решетка создает пространственно разделенные периодические возмущения различных периодов. Исключение составляют импульсы, которые за решеткой распространяются под углом $\vartheta = 0$. В этом случае разложения не будет. Эти периодические возмущения, вообще говоря, не синусоидальны. Но, как и всякие периодические функции, они могут быть разложены в ряд Фурье с основным периодом τ. Для понимания действия решетки такое разложение не обязательно. Но оно целесообразно, когда приемниками излучения в спектроскопах и спектрографах являются глаз, фотографическая пластинка и другие селективные приемники. Они действуют подобно набору гармонических осцилляторов с различными собственными частотами.

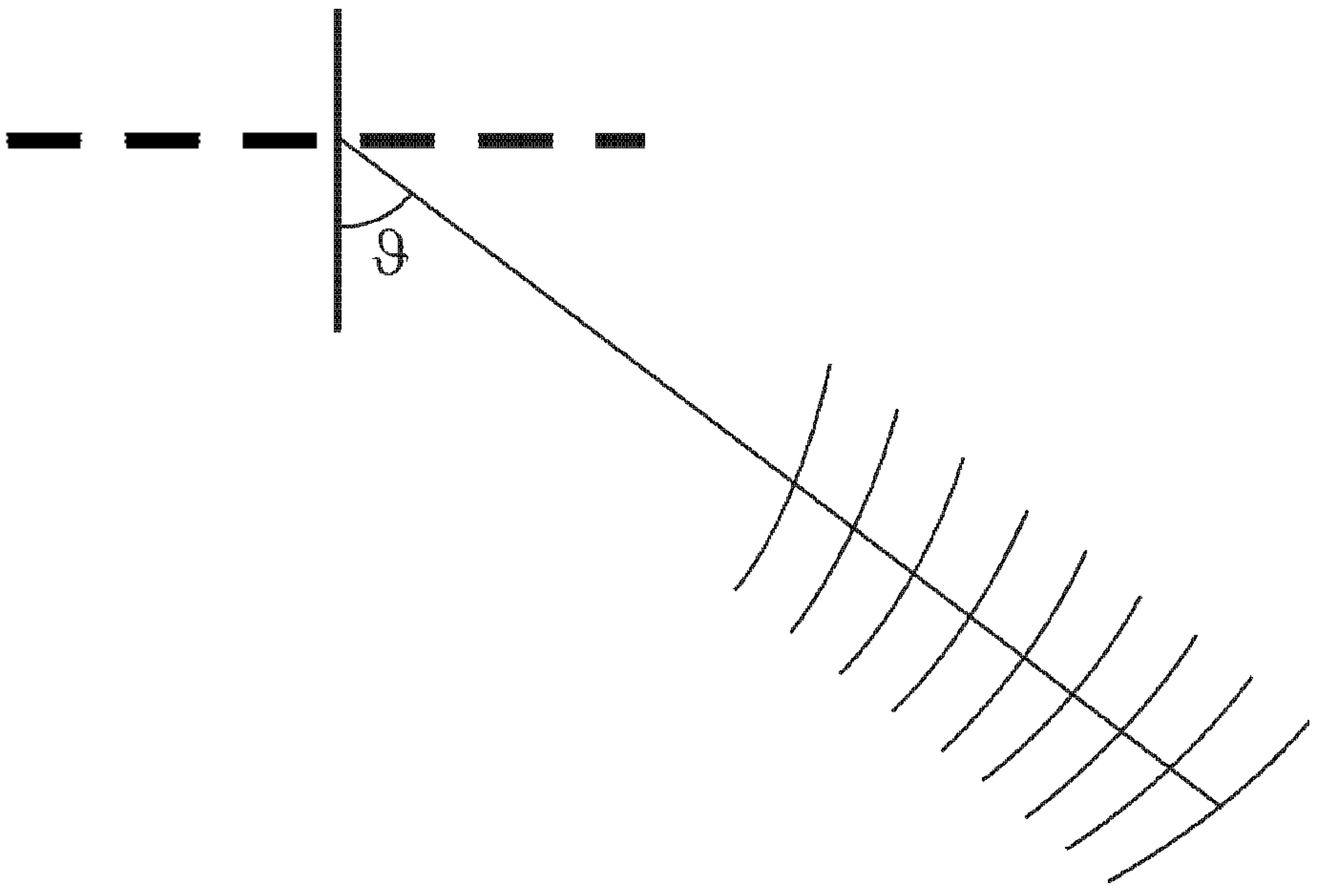


Рис. 200

Ряд Фурье, в который может быть разложено периодическое возмущение $E(t)$, вообще говоря, будет содержать монохроматическую волну с основным периодом τ. Наряду с ней могут войти гармоники с периодами $\tau/2$, $\tau/3$, ... Не все эти гармоники воспринимаются приемником. В простейшем случае только одна из них, например гармоника с периодом τ/m, действует на приемник. Обозначим ее период через T. Тогда $T = \dfrac{\tau}{m} = \dfrac{d\sin\vartheta}{cm}$, или

$$d\sin\vartheta = m\lambda. \qquad (50.2)$$

Это — основная формула дифракционной решетки. Но могут быть и такие случаи, когда разложение Фурье содержит две или больше гармоник, действующих на приемник. Тогда имеет место перекрытие спектров разных порядков. Это, однако, возможно только тогда, когда падающее излучение заполняет спектральный интервал, ширина которого превосходит дисперсионную область решетки.

Рассмотрим теперь воздействие на решетку единичной монохроматической волны с периодом T. Вторичные волны, выходящие в этом случае из щелей решетки, а потому и их суперпозиция будут монохроматическими с тем же периодом T. Значит, монохроматическая волна после прохождения через спектральный аппарат остается монохроматической без изменения периода. Отсюда и из принципа суперпозиции следует, что *в спектре решетки не могут появиться монохроматические составляющие с частотами, отсутствующими в спектре падающего излучения.* Если спектральная область, занимаемая падающим излучением, не превосходит дисперсионную область, перекрытия спектров различных порядков совсем не будет. При этом в случае неограниченной решетки в спектре каждого порядка каждому направлению луча будет соответствовать строго определенная длина волны λ. Наложения волн различных частот не будет. Это значит, что разрешающая способность неограниченной решетки бесконечно велика.

2. Допустим теперь, что длина решетки L и число штрихов N конечны. Тогда периодические возмущения, формирующиеся решеткой, будут состоять из N импульсов, следующих один за другим через время $\tau = (d\sin\vartheta)/c$. Время

$$\Theta = N\tau = \frac{Nd\sin\vartheta}{c} = \frac{L\sin\vartheta}{c} \qquad (50.3)$$

приблизительно равно времени от момента появления первого до момента появления N-го импульсов. Оно называется *временем затягивания импульса*, так как за решеткой вместо одиночного короткого импульса получается волновой процесс длительностью Θ. Время Θ называют также *временем установления спектрального аппарата*, так как по прошествии такого времени начинают действовать все N штрихов решетки. Волновое возмущение в точке наблюдения может быть представлено в виде

$$E(t) = f_1(t) + f_2(t) + \ldots + f_N(t),$$

где слагаемые в правой части представляют импульсы, приходящие от отдельных штрихов. Так как эти импульсы запаздывают один относительно другого на время τ, а в остальном совершенно тождественны, то $f_2(t) = f_1(t-\tau)$, $f_3(t) = f_1(t-2\tau)$ и т.д. Результирующее возмущение представится в виде

$$E(t) = f_1(t) + f_1(t-\tau) + \ldots + f_1[t-(N-1)\tau].$$

Разложим его в интеграл Фурье

$$E(t) = \int\limits_0^\infty A(\omega)e^{i\omega t}\,d\omega, \quad \text{где} \quad A(\omega) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} E(t)e^{-i\omega t}\,dt.$$

Подставив сюда выражение для $E(t)$ и учтя, что $f_1(t-\tau) = f_1(t)$, получим

$$A(\omega) = \frac{1}{2\pi}[1 + e^{-i\omega\tau} + \ldots + e^{-i\omega(N-1)\tau}]\int\limits_{-\infty}^{+\infty} f_1(t)e^{-i\omega t}\,dt =$$
$$= F(\omega)\,\frac{\sin(N\omega\tau/2)}{\omega\tau/2}, \qquad (50.4)$$

где введено обозначение

$$F(\omega) = \frac{1}{2\pi}\,e^{-i(N-1)\omega\tau/2}\int\limits_{-\infty}^{+\infty} f_1(t)e^{-i\omega t}\,dt.$$

Величина $|F(\omega)|^2\,d\omega$ с точностью до постоянного множителя представляет интенсивность падающего импульса, приходящуюся на интервал частот $(\omega, \omega + d\omega)$. Такой же смысл имеет величина $|A(\omega)|^2\,d\omega$ для излучения, выходящего из решетки под углом ϑ. Теперь дифрагированное излучение, распространяющееся от решетки в определенном направлении, состоит не из одной монохроматической волны, как было в случае неограниченной решетки, а непрерывно распределено по спектру частот. Главные максимумы в этом распределении приходятся на частоты, при которых знаменатель в выражении (50.4) обращается в нуль, т.е. $\omega\tau/2 = m\pi$. Это приводит к соотношению (50.2). Главные максимумы не появятся, если $F(\omega) = 0$, т.е. когда в падающем импульсе не представлено излучение соответствующей частоты. Ближайший минимум в распределении интенсивности по частотам приходится на частоту, определяемую условием $N\omega\tau/2 = Nm\tau + \pi$, так как при такой частоте числитель в формуле (50.4) обращается в нуль. Величина

$$\delta\omega = \frac{2\pi}{N\tau} = \frac{2\pi}{\Theta} \qquad (50.5)$$

определяет порядок ширины спектральной области, которую заполняют волны, выходящие из решетки в одном и том же направлении. По своему содержанию соотношение (50.5), очевидно, совпадает с ранее выведенной и подробно проанализированной формулой (29.8).

Допустим теперь, что возмущение, падающее на решетку, состоит из одной только неограниченной синусоиды. Тогда частота ω в формуле (50.4) будет фиксирована. Но угол ϑ может меняться, а формула (50.4) определит зависимость амплитуды дифрагированного света от этого угла. Так как $\omega\tau = (2\pi/\lambda)d\sin\vartheta$, то эта зависимость представится формулой

$$A(\vartheta) = F(\vartheta)\frac{\sin N\dfrac{\pi d\sin\vartheta}{\lambda}}{\sin\dfrac{\pi d\sin\vartheta}{\lambda}}, \tag{50.6}$$

тождественной с формулой (46.1). Поэтому отпадает надобность в дальнейших рассуждениях. Заметим только, что разрешающая способность R связана с временем установления Θ соотношением

$$R = \Theta/T. \tag{50.7}$$

3. По описанной схеме действует всякий спектральный прибор. Это совершенно очевидно для интерференционных спектральных приборов, но не столь очевидно для призмы. В случае призмы время распространения света от волнового фронта AB (рис. 201) до другого волнового фронта $A'B'$ одно и то же для всех лучей. Однако это заключение относится к *фазе волны* и из него не следует, что световой импульс придет в точку наблюдения целиком (что было бы, если бы вещество призмы не обладало дисперсией), а не разобьется на ряд следующих друг за другом тождественных импульсов. В диспергирующей среде импульсы, вышедшие одновременно от различных точек волнового фронта AB, придут в соответствующие точки волнового фронта $A'B'$ не одновременно. Импульс растянется во времени, и можно говорить о времени затягивания Θ.

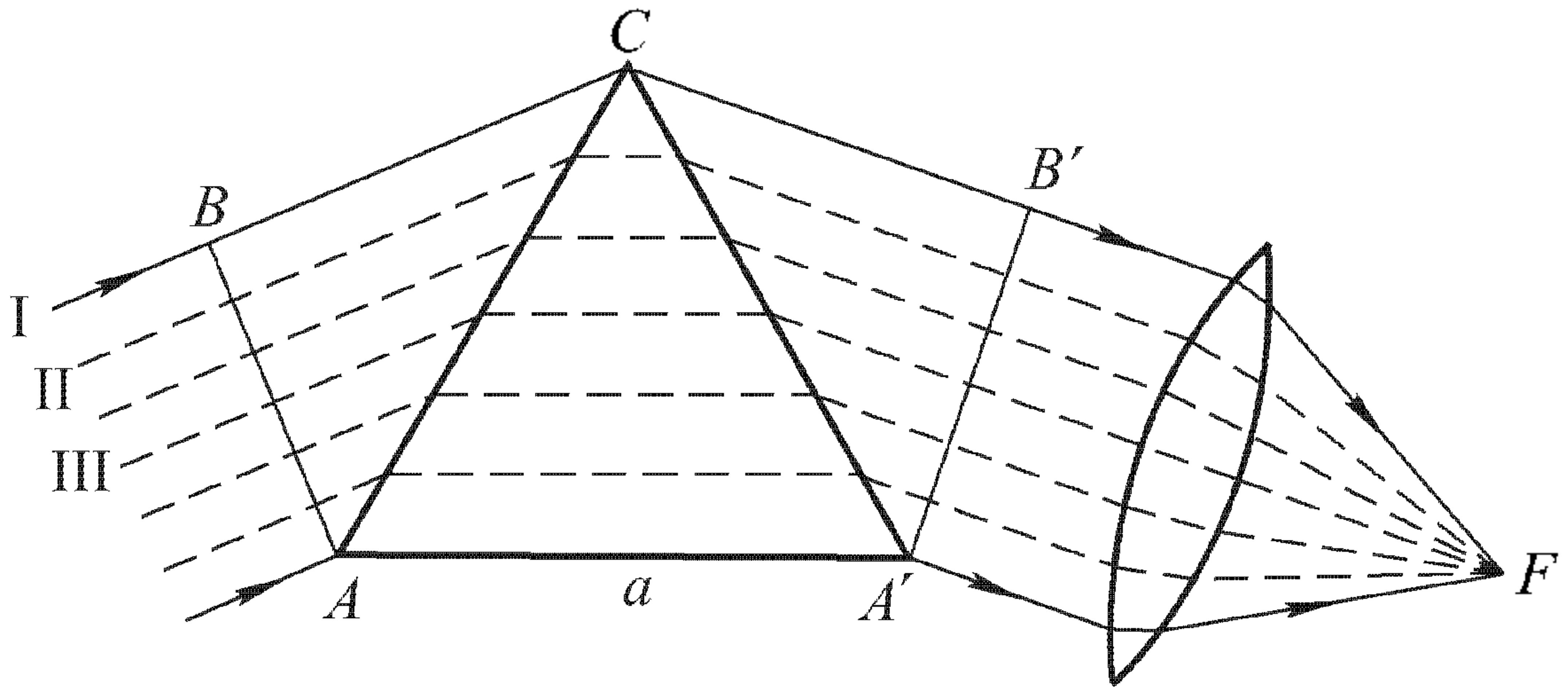


Рис. 201

Наиболее просто работу призмы можно проанализировать, когда падающий на нее импульс представляет группу волн, т.е. заполняет узкую область частот. Этого предположения достаточно для анализа наиболее важного вопроса — о разрешающей способности призмы. Как известно (см. § 8), при распространении в диспергирующей среде группа непрерывно деформируется, периодически восстанавливая свою форму через время восстановления $\tau = d\lambda/dv$. Разобьем падающий пучок на ряд пучков I, II, III, ... равноотстоящими плоскостями, параллельными лучам самого пучка. Разбиение произведем так, чтобы каждый пучок приходил в фокус F с запаздыванием τ по отношению к предыдущему пучку. Тогда волновые импульсы, связанные с рассматриваемыми пучками, по выходе из призмы будут иметь одну и ту же форму. Через фокус F последовательно будут проходить тождественные световые импульсы, следующие друг за другом через одно и то же время τ. Призма преобразовала падающий на нее импульс в последовательность равноотстоящих импульсов, т.е. в периодическое возмущение, а это и требуется от спектрального аппарата. Время затягивания, очевидно, равно $\Theta = AA'/u - BCB'/c$. По определению волнового фронта $BCB'/c = AA'/v = a/v$. Здесь v означает фазовую, а u — групповую скорости света в веществе призмы. Таким образом,

$$\Theta = a\left(\frac{1}{u} - \frac{1}{v}\right) = a\,\frac{v-u}{uv},$$

или на основании формулы (8.8)

$$\Theta = -a\,\frac{\dfrac{\lambda}{n}\dfrac{dn}{d\lambda}}{v\left(1+\dfrac{\lambda}{n}\dfrac{dn}{d\lambda}\right)} = -\frac{a}{v}\frac{\lambda\,dn}{d(n\lambda)}. \tag{50.8}$$

Поделив это выражение на T, найдем разрешающую способность призмы:

$$R = -a\,\frac{dn}{d(n\lambda)}. \tag{50.9}$$

Здесь λ означает длину волны в призме, тогда как в формуле (49.2) той же буквой обозначена длина волны в вакууме. Обозначив последнюю через λ_0, получим

$$R = -a\,\frac{dn}{d\lambda_0}, \tag{50.10}$$

что по содержанию совпадает с формулой (49.2). С изложенной точки зрения особенно ясно действие системы призм, разрешающая способность которой выражается формулой (49.3).

§ 51. Вогнутая отражательная решетка

1. Свет, дифрагировавший на обычной плоской решетке с равноотстоящими штрихами, должен направляться на линзу или другое фокусирующее устройство, чтобы в фокальной плоскости получилась картина резких спектральных линий. Однако можно обойтись и без линзы, если штрихи на плоской поверхности нанести неравномерно, надлежащим образом монотонно уменьшая или увеличивая расстояние между ними. Можно также придать поверхности решетки вогнутую сферическую форму и нанести на нее равноотстоящие штрихи.

Фокусирующее действие многих решеток было замечено давно, но до работ Роулэнда оно рассматривалось как недостаток решетки. Роулэнд превратил его в достоинство решетки, построив *вогнутые отражательные решетки*. Наряду с интерферометром Фабри–Перо такие решетки до настоящего времени являются основными спектральными аппаратами, широко используемыми в точных спектроскопических исследованиях. Спектральные линии, получаемые с помощью вогнутых отражательных решеток, по своей резкости значительно превосходят те же линии в решетках, где фокусировка осуществляется с помощью линз.

2. Пусть G_0G (рис. 202) — отражающая сферическая поверхность радиуса ρ с центром в точке C, на которой нанесены штрихи равноотстоящими плоскостями, перпендикулярными к плоскости рисунка и параллельными диаметру OC, а d — расстояние между этими плоскостями (период решетки). Поместив источник света в точке A_0, выясним, при каких условиях спектр m-го порядка можно сфокусировать в заданной точке A, лежащей в плоскости A_0OC. При исследовании ограничимся рассмотрением лучей, лежащих в той же плоскости, приняв ее за координатную плоскость XY. Начало координат поместим в центре решетки O, ось X направим вдоль диаметра OC, ось Y — по касательной к сфере, лежащей в рассматриваемой плоскости. Пусть $P(x, y)$ — точка на поверхности решетки, находящаяся у какого-либо штриха. Введем обозначения: $A_0O = r_0$, $AO = r$, $A_0P = u_0$, $AP = u$. Для фокусировки в точке A спектра m-го порядка необходимо и достаточно, чтобы разность хода между лучами, приходящими от соседних штрихов, составляла $m\lambda$. Предполагая (что не влияет на результат), что начало O находится также у штриха, запишем это условие в виде $u_0 + u = r_0 + r \pm nm\lambda$, где n — номер штриха P. При описанном выше способе нанесения штрихов $n = \pm y/d$, а потому при любом y должно быть

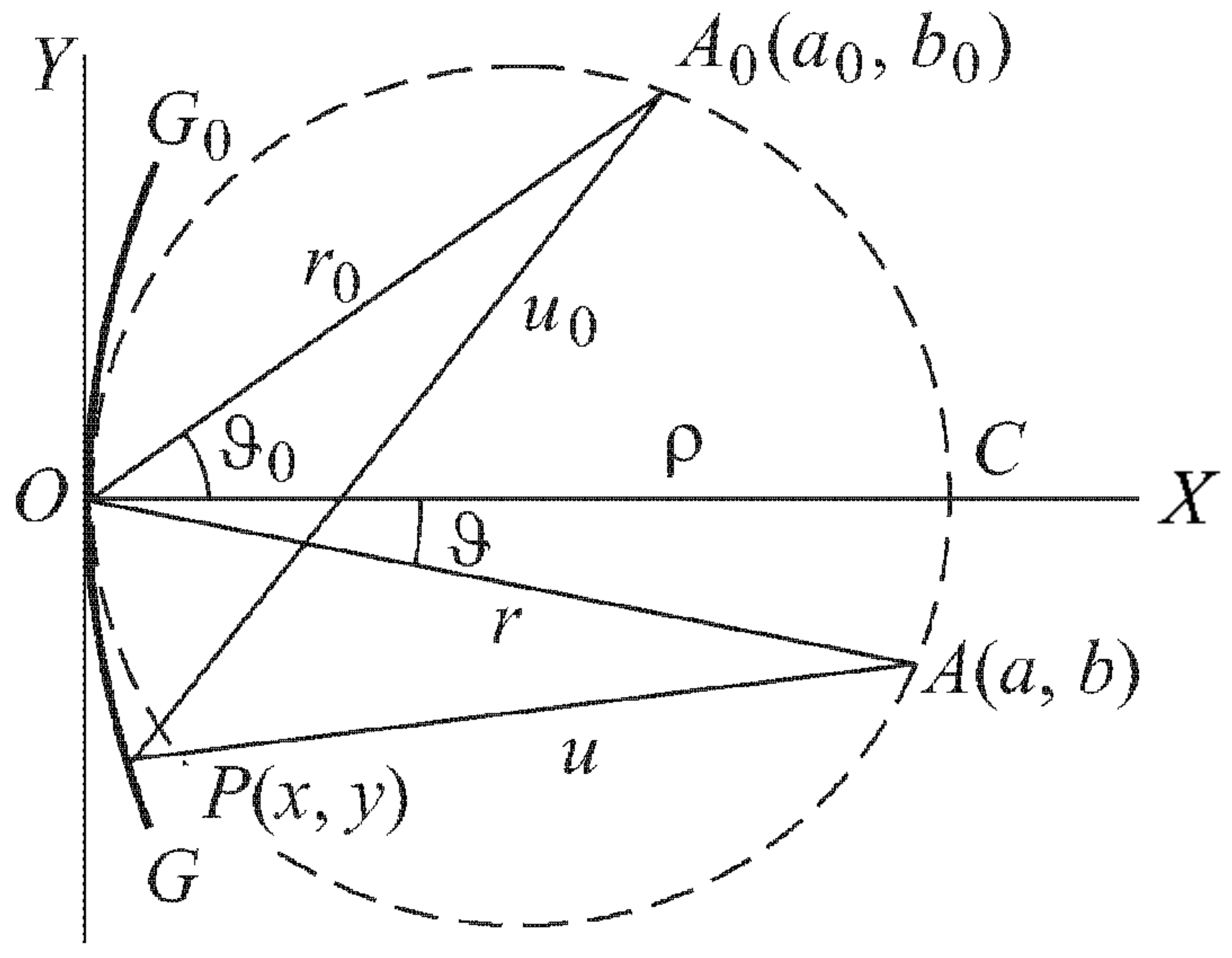


Рис. 202

$$u_0 + u = r_0 + r \pm \frac{m\lambda}{d} y. \tag{51.1}$$

Обозначим через a_0, b_0 координаты точки A_0; a, b — координаты точки A. Тогда

$$u_0^2 = (y - b_0)^2 + (x - a_0)^2 = r_0^2 + x^2 + y^2 - 2b_0y - 2a_0x. \tag{51.2}$$

Уравнение окружности, по которой поверхность решетки пересекается с плоскостью рисунка, имеет вид

$$x^2 + y^2 = 2\rho x. \tag{51.3}$$

Подставив отсюда $2x$ в предыдущее уравнение и использовав соотношение $r_0^2 - b_0^2 = a_0^2$, найдем

$$u_0^2 = \left(r_0 - \frac{b_0 y}{r_0}\right)^2 + \left(\frac{a_0}{r_0^2} - \frac{1}{\rho}\right) a_0 y^2 + \left(1 - \frac{a_0}{\rho}\right) x^2. \tag{51.4}$$

Второй член — величина второго, а последний — четвертого порядка малости по y. Сохраняя только члены второго порядка, получим

$$u_0 = r_0 - \frac{b_0 y}{r_0} + \frac{1}{2r_0}\left(\frac{a_0}{r_0^2} - \frac{1}{\rho}\right) a_0 y^2. \tag{51.5}$$

Аналогично,

$$u = r - \frac{by}{r} + \frac{1}{2r}\left(\frac{a}{r^2} - \frac{1}{\rho}\right) a y^2. \tag{51.6}$$

Подставим эти выражения в формулу (51.1) и приравняем нулю члены с первыми и вторыми степенями y. Тогда получим два условия:

$$\frac{b_0}{r_0} + \frac{b}{r} = \mp\frac{m\lambda}{d}, \tag{51.7}$$

$$\frac{a_0}{r_0}\left(\frac{a_0}{r_0^2} - \frac{1}{\rho}\right) + \frac{a}{r}\left(\frac{a}{r^2} - \frac{1}{\rho}\right) = 0, \tag{51.8}$$

при выполнении которых, по крайней мере с рассматриваемой степенью точности, решетка будет обладать требуемыми фокусирующими свойствами.

3. Условие (51.7) определяет *направления на дифракционные максимумы* m-го порядка. Действительно, если ϑ_0 — угол падения центрального луча на решетку, а ϑ — угол между соответствующим дифрагированным лучом и нормалью к ней (рис. 202), то $\sin\vartheta_0 = b_0/r_0$, $\sin\vartheta = b/r$, и уравнение (51.7) принимает вид

$$d(\sin\vartheta + \sin\vartheta_0) = \pm m\lambda, \tag{51.9}$$

такой же, что и при дифракции на плоской отражательной решетке.

Второе условие, (51.8), определяет расстояние r точки A от центра решетки O. Вместе с уравнением (51.9) оно, таким образом, однозначно определяет положение этой точки, если известно положение источника света A_0. Используя соотношения $a_0 = r_0\cos\vartheta_0$ и $a = r\cos\vartheta$, уравнение (51.8) можно представить в виде

$$\frac{\cos^2\vartheta_0}{r_0} + \frac{\cos^2\vartheta}{r} = \frac{\cos\vartheta_0 + \cos\vartheta}{\rho}. \tag{51.10}$$

При малых углах ϑ_0 и ϑ получается общеизвестное уравнение вогнутого зеркала радиуса ρ. Таким образом, каково бы ни было положение источника света A_0, всегда существует такая точка A, в которой приближенно собираются дифрагированные лучи. Однако можно достигнуть большей точности и простоты в установке решетки, а также лучшей резкости спектральных линий, если распорядиться выбором точки A_0 так, чтобы в выражении

(51.4) исчезли члены второго порядка. Для этого должно быть $\dfrac{a_0}{r_0^2} - \dfrac{1}{\rho} = 0$. Тогда, ввиду (51.8), будет $\dfrac{a}{r^2} - \dfrac{1}{\rho} = 0$. Эти соотношения можно записать в виде

$$r_0^2 = a_0\rho, \quad \text{или} \quad r_0 = \rho\cos\vartheta_0, \tag{51.11}$$

$$r^2 = a\rho, \quad \text{или} \quad r = \rho\cos\vartheta. \tag{51.12}$$

Соотношение (51.11) означает, что источник света должен лежать на окружности диаметра ρ, касающейся поверхности решетки в точке O. Из (51.12) тогда следует, что дифрагированные лучи соберутся в точке A, лежащей на той же окружности. Эта окружность называется *кругом Роулэнда*. На рис. 202 она изображена штриховой линией.

Угловая дисперсия решетки $d\vartheta/d\lambda = m/(d\cos\vartheta)$ не будет зависеть от угла ϑ, если спектры получаются при малых ϑ вблизи $\vartheta = 0$. Тогда $\cos\vartheta \approx 1$, угол ϑ будет линейной функцией длины волны, т.е. получится нормальный спектр. Поэтому в обычных условиях решетка устанавливается так, чтобы изучаемый спектр получался вблизи ее главной оптической оси.

Большие решетки в первом порядке имеют разрешающую способность около 200 000. Угловое расстояние между спектральными линиями, еще разрешаемыми решеткой, составляет $m\delta\lambda/(d\cos\vartheta) = m\lambda/(Rd\cos\vartheta)$, а линейное

$$\Delta x = \frac{m}{d\cos\vartheta}\frac{\rho\lambda}{R} \approx \frac{m}{d}\frac{\rho\lambda}{R}. \tag{51.13}$$

Для эффективного использования разрешающей способности решетки необходимо, чтобы величина Δx была не меньше определенного предела. Положим $\Delta x = 0{,}01$ мм, $m = 1$, $R = 200\,000$, $\lambda = 500$ нм, $d = 1/600$ мм. Тогда по предыдущей формуле найдем, что радиус кривизны решетки ρ должен быть не меньше ≈ 7 м, т.е. очень велик.

4. Вычисления, приведенные выше, применимы только к лучам, лежащим в плоскости рисунка. Более подробное исследование показывает, что лучи, выходящие из этой плоскости, дают *астигматические изображения*: точечный источник, помещенный в A_0, изображается в виде отрезка, перпендикулярного к плоскости рисунка и проходящего через точку A. Астигматизм увеличивается по мере возрастания угла падения ϑ_0, т.е. с увеличением порядка спектра. Когда источником света является щель, то наличие астигматизма ведет к уменьшению резкости изображения и понижению разрешающей способности, если щель установлена не совсем параллельно штрихам решетки. Поэтому необходимо добиваться тщательной установки щели на параллельность штрихам решетки. По той же причине нельзя делать штрихи решетки очень длинными. Обычно наносят короткие штрихи вдоль широких участков, а не стремятся покрыть ими большую площадь, как в случае плоской решетки, не обладающей астигматизмом.

5. Роулэнду принадлежит и теоретически наиболее совершенный способ установки дифракционной решетки, показанный на схематическом рис. 203. Решетка G_0G и фотографическая камера устанавливаются на противоположных концах твердого стержня OA, длина которого равна радиусу кривизны решетки ρ. Они могут скользить вдоль взаимно перпендикулярных рельсов A_0O и A_0A на салазках, шарнирно соединенных со стержнем OA. Для достижения лучшей отчетливости изображения фотопластинку P_0P изгибают по окружности, совпадающей с кругом Роулэнда. Щель помещается в точке пересечения A_0 направляющих рельсов. При фотографировании спектров разных порядков, а также различных участков одного и того же спектра щель A_0 и источник света L остаются неподвижными. Так как точки O, A_0, A лежат на круге Роулэнда, то при надлежащем положении стержня OA изображение щели A_0 получится в точке A. При перемещении концов стержня вдоль направляющих рельсов через точку A будут последовательно проходить фокусы различных длин волн в спектре рассматриваемого порядка. Так как в установке Роулэнда $\vartheta = 0$, то $d\sin\vartheta_0 = m\lambda$. С другой стороны, $OA_0 = \rho\sin\vartheta_0$. Комбинируя эти две формулы, находим

$$OA = \frac{\rho m}{d}\lambda. \qquad (51.14)$$

На стержне, связанном с рельсом OA_0, можно нанести равномерную шкалу длин волн. Тогда длина волны, попадающей в центр поля зрения A, может быть отсчитана непосредственно по этой шкале.

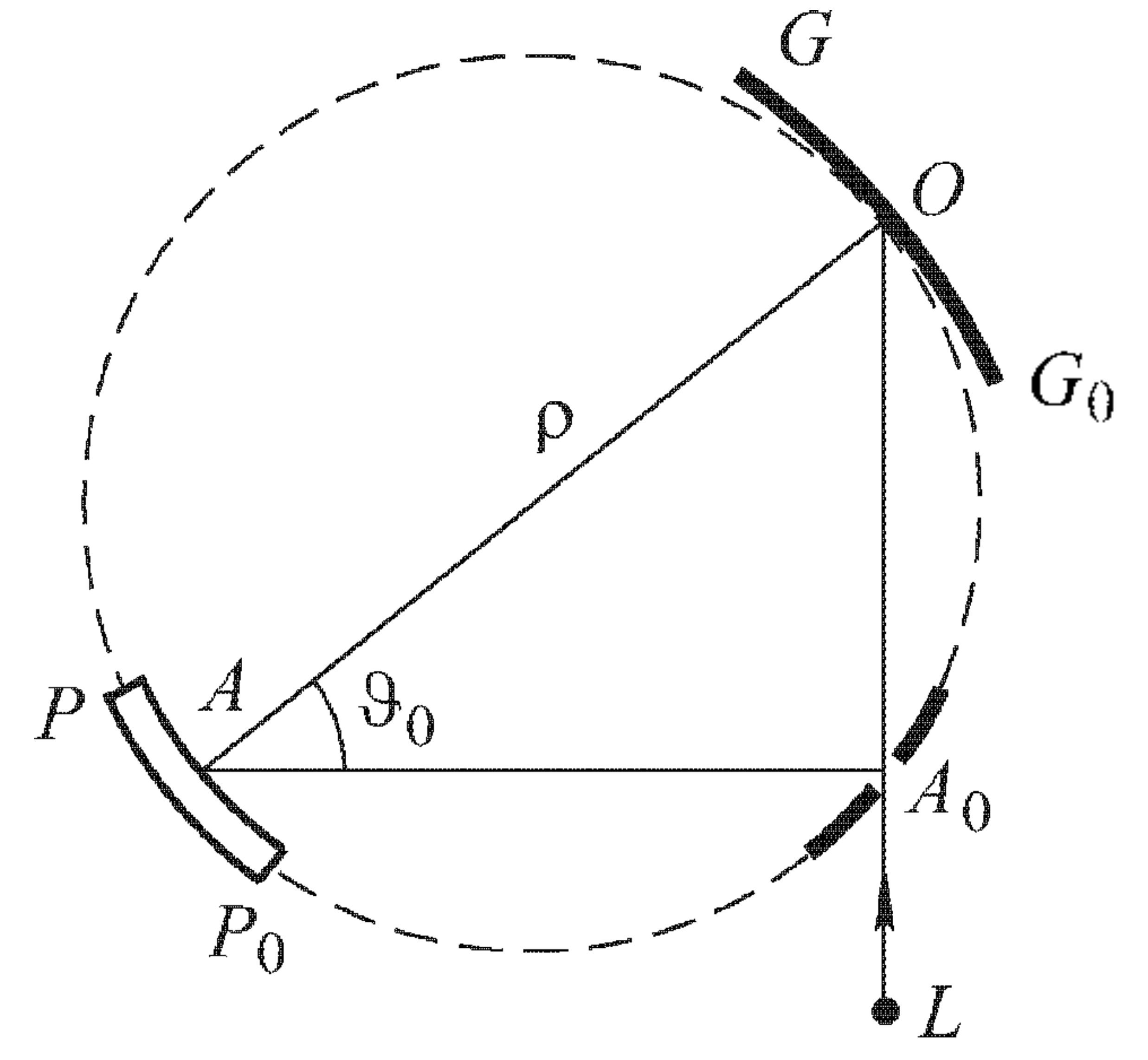


Рис. 203

При установке решетки по способу Роулэнда объем, занимаемый приборами, настолько велик, что в нем трудно поддерживать постоянство температуры, необходимое при длительных экспозициях. Кроме того, установка Роулэнда довольно дорога. По этим и по ряду других причин она применяется редко. Пашеном, Иглом и др. были разработаны другие способы; с ними можно ознакомиться по специальным руководствам по технике спектроскопии.

§ 52. Дифракция на решетке как краевая задача

1. В предыдущих параграфах волновое поле за решеткой вычислялось путем *суммирования волн*, исходящих от штрихов решетки. Для некоторых целей более предпочтителен другой способ. Допустим сначала, что решетка бесконечна. Переднюю поверхность ее будем называть *входом*, заднюю — *выходом*. Эта терминология применима и для отражательной решетки. Для нее входом и выходом служит одна и та же (передняя) поверхность. Без потери общности можно рассуждать так, как если бы решетка была бесконечно тонкой. Примем плоскость решетки за координатную плоскость XY. Ось X направим перпендикулярно к штрихам, а ось Z — в сторону распространения дифрагированного света (рис. 204). Как и раньше, отвлечемся от векторного характера волнового поля, считая его скалярным. Это всегда можно сделать, если электрический или магнитный вектор падающей волны параллелен штрихам решетки. Надо только в первом случае световое поле характеризовать электрическим, а во втором — магнитным вектором. В общем случае падающую волну можно разложить на две составляющие, из которых одна поляризована в плоскости падения, а другая перпендикулярно к ней. Тогда задача о дифракции волны сведется к двум независимым задачам, отличающимся состоянием поляризации падающего света.

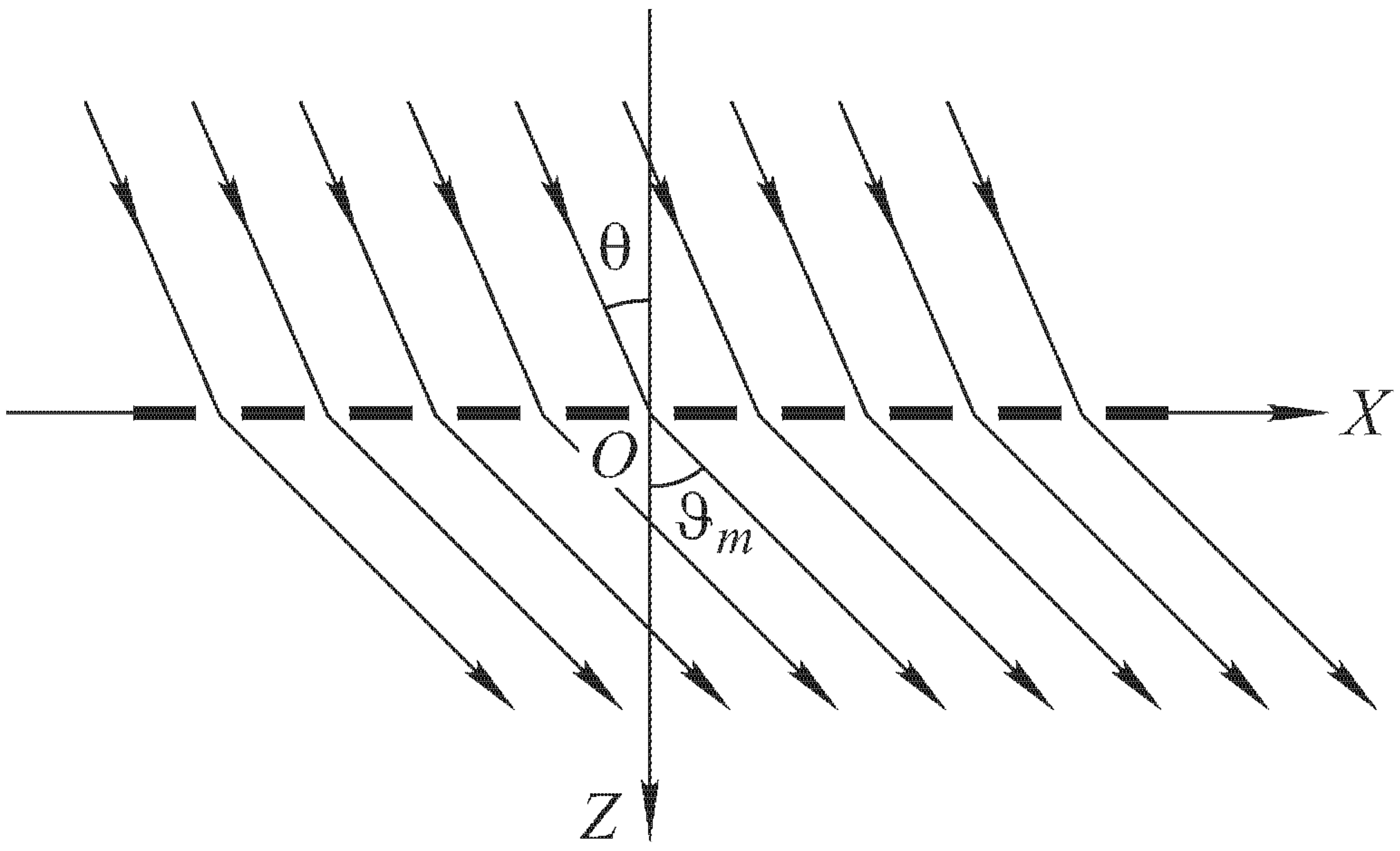


Рис. 204

Падающую волну представим в виде

$$E_0 = Ae^{i(\omega t - \mathbf{kr})}. \tag{52.1}$$

Полагая здесь $z = 0$, найдем поле $E_{\text{вх}}$ на входе решетки. В силу линейности уравнений, определяющих распространение волн, суммарное поле прошедшей (отраженной) и дифрагированных волн на выходе можно представить в виде $E_{\text{вых}} = DE_{\text{вх}}$. Коэффициент D, вообще говоря, зависит от длины волны и называется *пропускаемостью* или *амплитудной прозрачностью решетки*. Это есть характеристика *только самой решетки*. Для двумерной решетки величина $D = D(x, y)$ есть функция двух координат, для одномерной — одной координаты: $D = D(x)$.

Задача о дифракции на всякой решетке сводится к нахождению амплитудной прозрачности последней. Действительно, знание этой величины позволяет определить поле на выходе решетки, а этого достаточно, чтобы по принципу Гюйгенса–Френеля вычислить поле дифрагированной волны. Для нахождения самого решения, конечно, не обязательно пользоваться принципом Гюйгенса–Френеля. Достаточно найти решение волнового уравнения

$$\Delta E + k^2 E = 0, \tag{52.2}$$

которое в плоскости $z = 0$ переходило бы в функцию $E_{\text{вых}}(x, y) = D(x, y)E_{\text{вх}}(x, y)$ и удовлетворяло некоторым дополнительным требованиям, определяющимся физическими условиями задачи. Такой метод в литературе известен под названием *метода Рэлея*, хотя он и является только частным случаем общего *метода краевых задач* математической физики. Здесь этот метод применяется к одномерной дифракционной решетке.

2. Поскольку для решетки D есть периодическая функция x с периодом d, ее можно представить рядом Фурье

$$D = \sum_{m=-\infty}^{+\infty} D_m e^{-impx} \qquad \left(p = \frac{2\pi}{d}\right). \tag{52.3}$$

Поэтому

$$E_{\text{вых}} = DE_{\text{вх}} = \sum_{m=-\infty}^{m=+\infty} AD_m e^{i[\omega t-(k_x+mp)x]}. \tag{52.4}$$

На это соотношение надо смотреть как на *граничное условие*, которому должно удовлетворять за решеткой решение уравнения (52.2). Частным решением этого уравнения является плоская волна $E = ae^{i(\omega t-\mathbf{qr})}$, волновой вектор $\mathbf{q}$ которой удовлетворяет условию $\mathbf{q}^2 = k^2$. Он может быть и вещественным (однородная волна), и комплексным (неоднородная волна). Суперпозицией таких решений можно получить общее решение:

$$E = \sum_{m=-\infty}^{+\infty} a_m e^{i(\omega t-q_m\mathbf{r})}. \tag{52.5}$$

Чтобы при этом удовлетворить граничному условию (52.4), необходимо положить

$$q_{mx} = k_x + mp, \quad q_{my} = 0. \tag{52.6}$$

Тогда $q_{mz} = \sqrt{k^2 - q_{mx}^2}$. Двузначность квадратного корня устраняется физическими соображениями. Вещественным значениям q_{mz} соответствуют однородные волны, которые могут только *уходить от решетки*, а не приходить к ней. Этому условию удовлетворяют только положительные значения q_{mz}. Мнимым соответствуют неоднородные волны. При удалении от решетки они могут *только затухать*. Поэтому коэффициент при мнимой части должен быть отрицательным. Таким образом, соотношение (52.6) надо пополнить условиями

$$q_{mz} = +\sqrt{k^2 - q_{mx}^2} \quad \text{и} \quad q_{mz} = -i\sqrt{k^2 - q_{mx}^2}, \tag{52.6а}$$

в которых квадратные корни понимаются в *арифметическом смысле*. После этого волновые векторы $\mathbf{q}_m$ определяются однозначно, а для поля на выходе можно написать

$$E_{\text{вых}} = \sum_{-\infty}^{+\infty} a_m e^{i(\omega t-q_{mx}x)}.$$

Сравнение этого выражения с (52.4) дает

$$a_m = AD_m, \tag{52.7}$$

чем и завершается решение задачи.

3. Полагая в (52.6) $k_x = (2\pi/\lambda)\sin\theta$, $q_{mx} = (2\pi/\lambda)\sin\vartheta_m$, $p = 2\pi/d$, получим

$$d(\sin\vartheta_m - \sin\theta) = m\lambda, \tag{52.8}$$

т.е. основную формулу дифракционной решетки. Однако в полученном решении содержатся и новые результаты. Во-первых, весь спектр за решеткой состоит из *одних только главных максимумов*, никаких второстепенных максимумов нет. Это, конечно, объясняется тем, что полученные результаты относятся к неограниченной решетке. Во-вторых, в формулу (52.5), наряду с однородными, входят и *неоднородные волны*, а потому она определяет волновое поле *на любых расстояниях от решетки*, а не только в волновой зоне, как было в случае формулы (46.2). При нормальном падении света наивысший порядок однородных волн m определяется условием $m \leqslant d/\lambda$. При $m > d/\lambda$ составляющая q_{mz} мнимая. Соответствующая неоднородная волна экспоненциально затухает по закону $\exp(-\varkappa_m z)$, где

$$\varkappa_m = \sqrt{k^2 - q_{mz}^2} = \frac{2\pi}{d}\sqrt{m^2 - (d/\lambda)^2}. \tag{52.9}$$

На расстоянии $z = 1/\varkappa_m$ амплитуда волны убывает в e раз, ее квадрат — в e^2 раз. Наибольшее из этих расстояний порядка d. Поэтому при $z \gg d$ поле за решеткой состоит только из однородных волн.

В частности, если $d < \lambda$, то при нормальном падении поле вдали от решетки состоит только из одной плоской волны нулевого порядка ($m = 0$). Никаких боковых дифракционных максимумов не появляется. Исследование прошедшей или отраженной волны при $z \gg \dfrac{\lambda d}{2\pi\sqrt{\lambda^2 - d^2}}$, или практически на расстояниях $z \gg \lambda$, не даст никакой информации о структуре решетки. Рассматривая решетку в микроскоп, мы не обнаружим на ней штрихов. По той же причине свет, распространяющийся в прозрачной среде, например в кристалле, не отклоняется в стороны, хотя среда и построена из дискретных атомов или молекул — кристалл является периодической структурой и в этом смысле может рассматриваться как дифракционная решетка (фактически наблюдаемое рассеяние света вызывается неоднородностями среды и тепловыми флуктуациями). В оптическом отношении кристалл ведет себя как однородная сплошная среда. Причина этого в том, что период молекулярной структуры кристалла меньше длины волны. Но как только нарушится это условие, появятся дифракционные максимумы. Возникнет интерференционное рассеяние волн. Такой случай реализуется в рентгеновской области спектра.

4. В непосредственной близи от решетки волновое поле имеет очень сложный вид. На расстояниях $z \gg 1/\varkappa_m$ поле становится более простым, поскольку из него выпадают неоднородные волны. Однако картина поля все еще довольно сложна, так как однородные волны, из которых оно состоит, накладываются друг на друга. При удалении от решетки на некоторое минимальное расстояние плоские волны различных порядков пространственно разделяются. На сравнительно малых расстояниях от решетки к пространственно разделенным волнам различных порядков можно применять геометрическую оптику. Направления на главные максимумы излучения определяются требованием, чтобы все волны, исходящие из различных щелей решетки, при интерференции усиливали друг друга. При этом явно или неявно всегда имеется в виду, что интерференция происходит либо в фокусе собирающей линзы, либо на бесконечном расстоянии от решетки. Однако поскольку дифрагированное поле всюду представляется суперпозицией плоских волн, условие интерференционного усиления определяет также направление распространения таких плоских волн на любых расстояниях от решетки.

5. Если в разложении пропускаемости D в ряд Фурье отсутствует какой-либо член, то будет отсутствовать и главный максимум соответствующего порядка в волновом поле за решеткой. Рассмотрим в качестве примера синусоидальную решетку, называемую также решеткой Рэлея. Ее пропускаемость определяется выражением

$$D = \alpha + \beta \cos px = \alpha + \frac{\beta}{2}\left(e^{ipx} + e^{-ipx}\right)$$

с постоянными a, β и p. В этом случае световое поле за решеткой состоит из трех плоских однородных волн: неотклоненной волны (максимум нулевого порядка) с амплитудой $a_0 = \alpha A$ и двух боковых максимумов первого порядка с амплитудами $a_1 = a_{-1} = {}^1\!/_2\beta A$. Других главных максимумов не будет. Если поверхности тела придать форму синусоиды, амплитуда которой мала по сравнению с длиной волны, то такая поверхность приближенно будет вести себя как решетка Рэлея. Это удалось сделать Вуду путем травления (примерно в течение 10 с) поверхности стеклянной дифракционной решетки слабым раствором фтористоводородной кислоты. Обычные стеклянные решетки дают очень слабые спектры, так как штрихи, нанесенные алмазом, узки по сравнению со светлыми промежутками между ними. Если штрихи расширить, смочив поверхность решетки слабой фтористоводородной кислотой, то удается достигнуть увеличения интенсивности спектров в несколько раз (до десяти).

Возможны и такие структуры, которые формально могут рассматриваться как дифракционные решетки, дающие только один максимум. Примером может служить призма. Если на одну из боковых граней призмы падает плоская монохроматическая волна, то фаза волнового поля на выходе (на второй грани призмы) будет меняться с координатой по линейному закону. Соответствующее граничное условие удовлетворяется только одной уходящей плоской волной. Это и есть волна, прошедшая через призму и отклоненная ею. Проблема отражения и преломления света на плоской границе раздела двух сред может решаться как краевая задача по методу Рэлея. При таком подходе плоская граница, раздела ведет себя как дифракционная решетка, дающая только один максимум.

6. Ограниченная решетка может быть рассмотрена также методом Рэлея. В пределах решетки ее пропускаемость периодична, а вне обращается в нуль. Разложение пропускаемости в интеграл Фурье имеет вид

$$D(x) = \int\limits_{-\infty}^{+\infty} C(f)e^{-\imath f x}\,df. \tag{52.10}$$

Поле на выходе решетки представится выражением

$$E_{\text{вых}} = A \int\limits_{-\infty}^{+\infty} C(f)e^{-f(k_x+f)x}\,df.$$

Решение, удовлетворяющее этому и всем прочим условиям задачи, имеет вид

$$E = A \int\limits_{-\infty}^{+\infty} C(f)e^{-\imath\mathbf{q}\mathbf{r}}\,df, \tag{52.11}$$

где $q_x = k_x + f$, $q_y = 0$, а q_z определяется прежними формулами (52.6а). Зададим теперь пропускаемость решетки выражением

$$D(x) = \begin{cases} 0, & \text{если} -\infty < x < -L, \\ \sum\limits_{m=-\infty}^{+\infty} D_m e^{-impx}, & \text{если} - L < x < +L, \\ 0, & \text{если} + L < x < +\infty. \end{cases} \tag{52.12}$$

Вычисляя коэффициент Фурье $C(f)$, получим

$$C(f) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} D(x)e^{ifx}\,dx = \frac{1}{\pi} \sum_{m=-\infty}^{+\infty} D_m \frac{\sin\left[L(f - mp)\right]}{f - mp}.$$

В результате поле за решеткой представится в вида

$$E = \sum E_m, \tag{52.13}$$

где

$$E_m = \frac{A}{\pi}\int\limits_{-\infty}^{+\infty} D_m \frac{\sin[L(f-mp)]}{f-mp}\, e^{-i\mathbf{qr}}\, df =$$

$$= \frac{Ad}{\pi}\int\limits_{-\infty}^{+\infty} D_m \frac{\sin\left[(N/a)(fd-2m\pi)\right]}{fd-2m\pi}\, e^{-i\mathbf{qr}}\, df. \qquad (52.14)$$

По-прежнему волновое поле за решеткой складывается из волн различных порядков m. Однако волна E_m уже не сводится к единственной плоской волне определенного направления, а состоит из суперпозиции *непрерывного множества* плоских волн различных направлений. Распределение интенсивности по направлениям в отдельной волне E_m определяется той же кривой, что и в волне, дифрагированной на щели (см. рис. 175). Интенсивность достигает максимума, когда знаменатель обращается в нуль, т.е. когда $q_x = k_x + mp = 0$. Это условие определяет направление на главный максимум m-го порядка. В рассматриваемом случае

$$\frac{\sin\left[(N/2)(fd-2m\pi)\right]}{fd-2m\pi} = \frac{N}{2},$$

а потому интенсивность волны E_m в главном максимуме пропорциональна N^2. Когда $\sin\left[(N/2)(fd-2m\pi)\right] = 0$, но $fd-2m\pi \neq 0$, интенсивность волны обращается в нуль. Этим определяются направления на дифракционные минимумы, число которых между двумя соседними главными максимумами равно $N-1$. Таким образом, получаются все результаты, найденные выше методом суммирования волн.

§ 53. Примеры на применение метода Рэлея

Ниже предполагается, что период решетки значительно больше длины световой волны, так что пропускаемость $D(x)$ можно вычислить в приближении геометрической оптики.

1. Начнем с плоской амплитудной решетки, состоящей из прозрачных щелей ширины b и непрозрачных промежутков между ними ширины a. В приближении геометрической оптики пропускаемость $D(x)$ равна единице на щели и нулю в промежутках между ними. Начало координат поместим в середине щели. Тогда коэффициент Фурье D_m представится выражением

$$D_m = \frac{1}{d}\int\limits_{-d/2}^{+d/2} D(x)e^{impx}\,dx = \frac{1}{d}\int\limits_{-b/2}^{+b/2} e^{impx}\,dx = \frac{b}{d}\frac{\sin(\pi mb/d)}{\pi mb/d}. \qquad (53.1)$$

При вычислении интенсивности дифрагированного пучка в принципе следовало бы учесть изменение его поперечного сечения из-за наклона к плоскости решетки. Однако это было бы превышением точности, так как наши вычисления применимы лишь при малых углах дифракции, когда $\cos\vartheta \approx 1$. В этом приближении относительная интенсивность m-го дифракционного пучка $I_m = D_m^2$. (Интенсивность падающей волны принята за единицу.) Для спектра нулевого порядка $D_0 = b/d$, $I_0 = (b/d)^2$. Полная интенсивность, пропускаемая решеткой, $I_{\text{прош}} = b/d$. Разность этих величин дает суммарную интенсивность света, приходящуюся на спектры прочих порядков:

$$\frac{b}{d} = \frac{b^2}{d^2} + \frac{2}{\pi^2} \sum_{m=1}^{\infty} \frac{1}{m^2} \sin^2 \frac{\pi m b}{d}. \tag{53.2}$$

(Это — чисто математическое соотношение, доказываемое в теории рядов Фурье.)

Относительная доля дифрагированного света будет

$$\frac{I_{\text{прош}} - I_0}{I_{\text{прош}}} = 1 - \frac{b}{d}. \tag{53.3}$$

Она максимальна и стремится к единице, когда $b/d \to 0$. Однако в этом случае сама интенсивность $I_{\text{прош}}$ также стремится к нулю. Интенсивность m-го дифракционного пучка равна

$$I_m = D_m^2 = I_0 \left[\frac{\sin(\pi m b/d)}{\pi m b/d}\right]^2. \tag{53.4}$$

Когда $b/d \to 0$, интенсивности всех дифракционных пучков становятся одинаковыми и равными I_0. Однако, как уже отмечено выше, в этом случае каждая из этих интенслвностей сама стремится к нулю.

Если $\pi m b/d = n\pi$, т.е. $b/d = n/m$, где n — целое число, меньшее m и взаимно простое с ним, то обращаются в нуль интенсивности спектров с порядками m, $2m$, $3m$, ... Так, при $b/d = 1/2$ пропадают все спектры четных порядков. Смысл этого результата, как уже отмечалось в § 46, станет очевидным, если заметить, что условие m-го главного максимума $d \sin\vartheta = m\lambda$ умножением на $b/d = n/m$ преобразуется в $b \sin\vartheta = n\lambda$, т.е. в условие n-го дифракционного минимума при дифракции на отдельной щели. Таким образом, под углом ϑ каждая щель, а потому и решетка в целом света не посылают.

2. Рассмотрим более общий случай. Допустим, что на участках длины b пропускаемость решетки равна β, а на участках длины a она равна α. Величины α и β постоянны, но могут быть комплексными. Таким образом, решетка является *амплитудно-фазовой*. Когда α и β — числа вещественные, то решетка будет *амплитудной*. Если же они — числа вида $e^{i\rho}$ (ρ вещественно), то решетка становится чисто *фазовой*. Рассматриваемая амплитудно-фазовая решетка эквивалентна плоскопараллельной пластинке с пропускаемостью α и наложенной на нее дифракционной решетке. Пропускаемость последней на участках b равна $(\beta - \alpha)$, а на участках a — нулю. Разумеется, величины α и β, a и b можно поменять местами и получить вторую эквивалентную систему. Математически обе системы отличаются одна от другой только обозначениями, а потому достаточно рассмотреть лишь одну из них, например, первую.

Вычисление коэффициентов Фурье D_m сводится к предыдущей задаче. Для плоскопараллельной пластинки все коэффициенты Фурье обращаются в нуль, за исключением нулевого, который равен α. Поэтому, поместив начало координат в центре одного из отрезков b и воспользовавшись формулой (53.1), получим

$$D_m = (\beta - \alpha)\,\frac{b}{d}\,\frac{\sin(\pi m b/d)}{\pi m b/d} + \alpha\delta_m, \tag{53.5}$$

где $\delta_m = 1$ при $m = 0$ и $\delta_m = 0$ при $m \neq 0$. При $\alpha = 0$, $\beta = 1$ получаются результаты предыдущей задачи.

3. Рассмотрим теперь частные случаи чисто амплитудной и чисто фазовой решеток. Для амплитудной решетки величины α и β вещественны и положительны. Все коэффициенты D_m также вещественны. Знаки этих коэффициентов, начиная с $m = \pm 1$, чередуются. Коэффициенты нулевого и первого порядков могут иметь одинаковые или противоположные знаки в зависимости от соотношения между пропускаемостями α и β. В случае чисто фазовой решетки пропускасмости α и β имеют вид $e^{i\rho}$. Так как существенна только разность фаз между волнами, исходящими из участков a и b, то без ущерба для общности можно положить $\alpha = 1$, $\beta = e^{i\rho}$. Тогда из формулы (53.5) находим

$$D_m = (e^{i\rho} - 1)\,\frac{b}{d}\,\frac{\sin{(m\pi b/d)}}{m\pi b/d} \qquad (m \neq 0), \tag{53.6}$$

$$D_0 = (e^{i\rho} - 1)\,\frac{b}{d} + 1. \tag{53.7}$$

Как и в случае амплитудной решетки, коэффициенты Фурье D_m, начиная с $m = \pm 1$, попеременно меняют знаки. Никакого дополнительного сдвига фаз между этими коэффициентами нет.

Качественное отличие фазовой решетки от амплитудной состоит в том, что в случае фазовой решетки имеется дополнительный сдвиг фаз φ между спектром нулевого и спектрами всех прочих порядков. Чтобы его вычислить, найдем из формул (53.6) и (53.7) комплексное отношение D_m/D_0. Аргумент этого комплексного числа и будет φ. Простое вычисление дает

$$\operatorname{tg}\varphi = \frac{b+a}{b-a}\,\frac{\sin\rho}{1-\cos\rho}. \tag{53.8}$$

Сдвиг фаз φ один и тот же для всех порядков m. Так как после дифракции на решетке спектры различных порядков пространственно разделяются на независимые пучки, то можно оказывать воздействие на каждый из них, не меняя при этом амплитуды и фазы всех остальных пучков. Например, если на пути нулевого пучка поставить прозрачную пластинку, которая изменила бы его фазу на φ, то фазовые соотношения между дифрагированными пучками будут такими же, как и у амплитудной решетки. С введением такой пластинки фазовая решетка действует как амплитудная. На этом основан *метод фазового контраста*, используемый в микроскопии (см. § 59).

Отметим два частных случая. Во-первых, случай $a = b$. Тогда формула (53.8) дает $\operatorname{tg}\varphi = \infty$, т.е. $\varphi = \pi/2$. Во-вторых, случай малых значений ρ. Тогда

$$\operatorname{tg}\varphi \approx \frac{b+a}{b-a}\,\frac{2}{\rho},$$

т.е. $\operatorname{tg}\varphi$ очень велик, а самый угол φ практически равен $\pi/2$. В обоих случаях для превращения фазовой решетки в амплитудную на пути нулевого пучка или на пути всех прочих дифрагированных пучков достаточно ввести пластинку, вносящую дополнительную разность фаз $\pm\pi/2$.

4. В качестве последнего примера рассмотрим фазовую решетку, профиль штрихов которой показан на рис. 205. Поперечное сечение штриха имеет форму треугольника, одна сторона которого длинная и пологая, а другая — короткая и крутая. Такая решетка интересна тем, что при определенных условиях она может концентрировать основную часть света в спектре одного порядка. Решетка может действовать и как пропускающая, и как отражательная. Ради определенности разберем действие пропускающей решетки.

Поместим начало координат O в середине длинной стороны AB. Пусть a — длина проекции длинной, а b — короткой стороны на направление оси X. Если $b \ll a$, то в интеграле (53.1) можно пренебречь вкладом, вносимым короткой стороной. В этом приближении можно положить $a = d$ и вычислить пропускаемость решетки только на участке a. Влияние пропускаемости участка b практически не отразится на результатах. Пусть волна падает перпендикулярно к плоскости XY и в воздухе представляется выражением $E = E_0 e^{-ikz}$. На входе, т.е. в плоскости $z = -h$, поле представляется выражением $E_{\text{вх}} = E_0 e^{ikh}$. Чтобы вычислить поле на выходе при $z = +h$, можно, ввиду малости угла наклона α, пренебречь преломлением. Если z_0 — текущая координата точки на прямой AB, то поле на выходе в лежащей под ней точкой будет равно

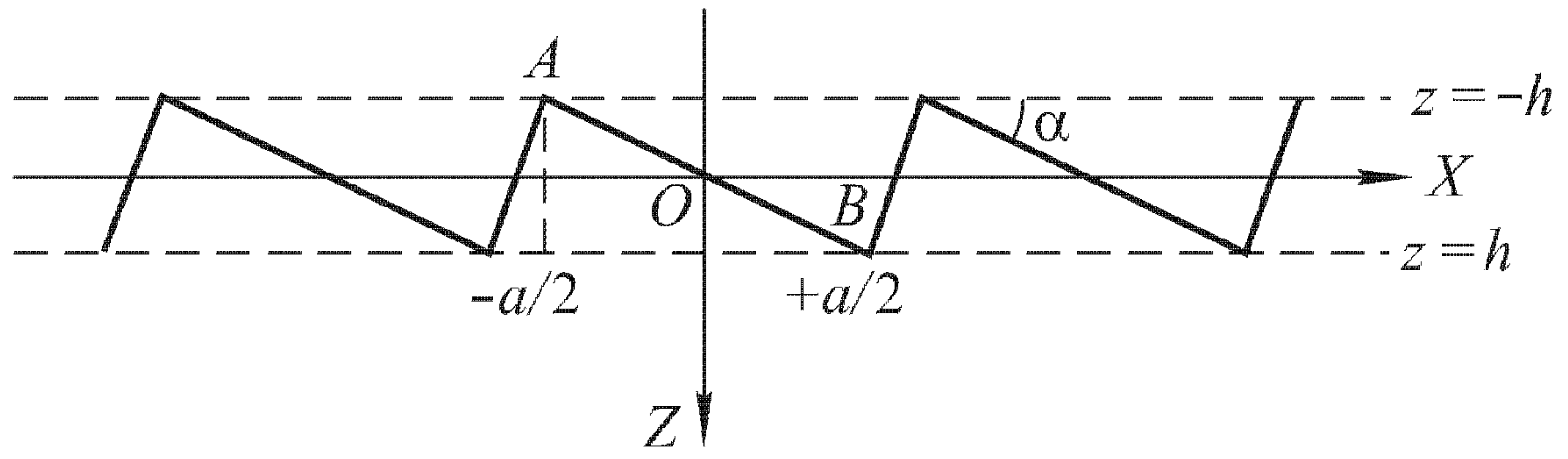


Рис. 205

$$E_{\text{вых}} = E_0 e^{-ikz_0} \cdot e^{-ikn(h-z_0)} = e^{i\delta} e^{ik(n-1)z_0} E_{\text{вх}},$$

где δ — некоторая постоянная. Постоянный фазовый множитель $e^{i\delta}$ не играет роли и может быть отброшен. Таким образом, пропускаемость решетки $D = e^{ik(n-1)z_0}$, или после подстановки $z_0 = x \operatorname{tg} \alpha \approx \alpha x$

$$D(x) = e^{ik(n-1)\alpha x}.$$

Коэффициенты Фурье D_m вычисляем по формуле (53.1) и находим

$$D_m = \frac{\sin \pi[m + (d/\lambda)(n-1)\alpha]}{\pi[m + (d/\lambda)(n-1)\alpha]}. \tag{53.9}$$

Если знаменатель этого выражения обращается в нуль, то почти весь свет сконцентрируется в спектре порядка m. Для этого должно быть $m + (d/\lambda)(n-1)\alpha = 0$. С другой стороны, $d \sin \vartheta = m\lambda$, или, ввиду малости угла дифракции, $d \cdot \vartheta = m\lambda$. Исключая m, получаем

$$\vartheta = -(n-1)\alpha. \tag{53.10}$$

Эта формула показывает, что угол ϑ равен углу отклонения луча при преломлении в призме с малым преломляющим углом α. Таким образом, почти весь свет может сконцентрироваться в одном направлении, если это направление совпадает с направлением преломленных лучей. Для концентрации необходимо, чтобы разность хода между пучками, преломленными на соседних ступеньках решетки, составляла целое число волн. С подобной концентрацией дифрагированного света в спектре одного порядка мы столкнулись также в § 48 при изучении эшелона Майкельсона.

§ 54. Голография

1. Применим метод Рэлея для уяснения идеи *голографии*, т.е. безлинзового получения оптических изображений путем так называемого *восстановления волнового фронта*. В принципе идея голографии была выдвинута и экспериментально проверена польским физиком М. Вольфке (1883–1947). Его работа была опубликована еще в 1920 г., но была забыта. Эту идею независимо от Вольфке вновь предложил и обосновал в 1947 г. английский инженер и физик Габор (р. 1900), который по праву считается изобретателем голографии. Однако понадобилось 15 лет, чтобы стало возможно практическое осуществление голографии. Причина столь длительной задержки заключается в том, что в голографии требуются источники света, обладающие *высокой степенью временной и пространственной когерентности*. Таких источников в 1947 г. еще не существовало. Положение изменилось в 1960 г. с изобретением лазеров и проникновением их в лабораторную технику. Первые изображения по методу голографии были получены американцами Лейтом и Упатниексом в 1962 г.

При освещении или просвечивании предмета от него распространяется рассеянная или прошедшая волна. (Для определенности в дальнейшем предполагается, что предмет освещается, а не просвечивается, хотя принципиального различия между этими случаями при нашем рассмотрении нет.) Отделившись от предмета, рассеянная волна сохраняет в дальнейшем *независимое существование* и несет полную информацию о форме и прочих свойствах предмета, какая может быть получена путем освещения его световыми лучами. Попадая в глаз или объектив фотоаппарата, эта волна образует на сетчатке или на фотопластинке изображение предмета. Если любым путем создать такую же волну, то, очевидно, она сможет вызвать в точности такие же эффекты, что и исходная волна, рассеянная предметом. На этом замечании и основана идея голографии.

Процесс получения изображения в голографии распадается на две стадии. На первой стадии изготовляется *голограмма*, т.е. фотопластинка, с помощью которой можно восстанавливать световую волну, рассеянную телом. Вторую стадию составляет само *восстановление* этой волны и получение оптического изображения.

2. Пусть какой-то предмет A (рис. 206) освещается пучком паралелльных лучей от лазера. Рассеянные лучи попадают на фотопластинку Γ. По степени почернения пластинки после проявления можно судить об амплитуде рассеянной волны во всех местах пластинки, которых эта волна достигла. В этом смысле экспонированная и проявленная пластинка сохраняет информацию об амплитуде волнового поля. Для восстановления волнового поля такой информации, конечно, недостаточно. Нужна еще дополнительная информация о *фазе*, которой пластинка не содержит, так как степень почернения зависит только от интенсивности, но не от фазы волны. Габор указал, что необходимую информацию о фазе можно получить и записать на той же фотопластинке Γ, если осветить ее вторым пучком от того же лазера и заставить его интерферировать с пучком, рассеянным предметом. Практически этого можно достигнуть, расширив предварительно пучок от лазера, а затем разделив его на два пучка. Один из них (*предметный*) направляется на предмет A, другой (*опорный*) отражается от плоского зеркала S. Оба пучка направляются на фотопластинку Γ и там интерферируют между собой. Интерференционная картина фотографируется. Так полученная фотография и называется *голограммой*.

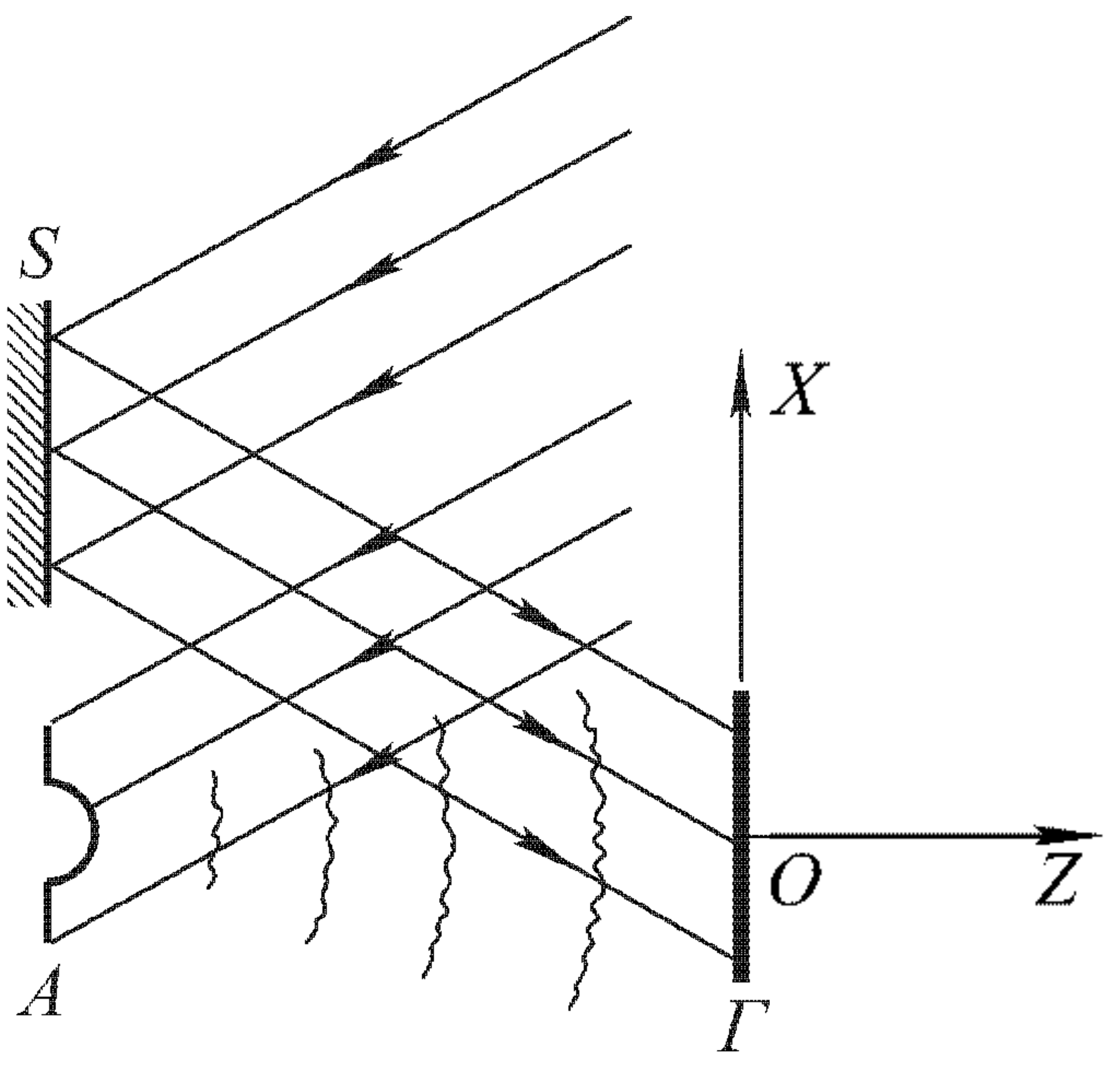


Рис. 206

Поскольку волна, рассеянная предметом, возникает при отражении и дифракции на *макроскопических* деталях предмета со сложной формой и взаимным расположением, реальная голограмма представляет собой очень сложную и запутанную интерференционную картину с очень мелкими деталями, которые невозможно различить невооруженным глазом. (На ней обычно видны и крупные дифракционные кольца. Но они не имеют никакого отношения к делу, так как возникают при дифракции на случайных пылинках, встретившихся на пути распространения света.)

Дифракционная картина на голограмме не имеет ни малейшего сходства с предметом. При рассматривании ее в микроскоп в ней трудно усмотреть следы каких-либо закономерностей. И тем не менее расположение, форма и интенсивность дифракционных пятен голограммы полностью определяются геометрической формой и физическими свойствами отражающей поверхности объекта. Голограмма *в закодированной форме* содержит *полную информацию* об амплитудах и фазах рассеянной волны, которая достаточна для ее восстановления и получения оптического изображения. Само название «голография» происходит от греческих слов «голог» — полный и «графе» — пишу и может быть переведено как «полная запись».

3. Разность хода между опорной волной и волнами, рассеянными предметом, очень велика и может составлять несколько метров. Поэтому при изготовлении голограммы время когерентности света должно быть очень велико (не меньше 10^{-5}–10^{-7} с). Длина когерентности также должна быть очень большой (не меньше 1–10 м). Никакие источники света, кроме лазеров, не могут обеспечить столь высокую степень временной и пространственной когерентности. Кроме того, необходима *неподвижность* (с точностью до долей длины световой волны) и *высокая контрастность* интерференционной картины, образующейся в области перекрытия предметного и опорного пучков во время экспозиции. Для этого также требуется высокая когерентность света, а также *механическая жесткость* всех элементов голографической установки.

Необходимая степень монохроматичности света определяется по общей формуле $\lambda/\delta l \gtrsim m$, где m — максимальный порядок интерференции, наблюдающейся при голографировании. При рациональном расположении источника света и прочей аппаратуры его можно оценить по формуле $m \sim L/\lambda$, где L — линейные размеры предмета. Таким образом, должно быть $\delta\lambda < \lambda^2/L$. Уже при небольших размерах предмета ~ 10 см и $\lambda = 500$ нм эта формула дает $\delta\lambda < 10^{-3}$ нм. Между тем спектральные линии ртутной лампы, даже с низким давлением ртутных паров, имеют ширину ~ 30 нм. Требования к размерам источника света, пожалуй, еще более жесткие. Для получения высокой контрастности интерференционной картины на голограмме поперечные размеры Δx источника света (т.е. размеры параллельно плоскости голограммы) должны быть малы по сравнению с шириной интерференционной полосы λ/α, где α — угол схождения крайних интерферирующих лучей (см. § 28). По порядку величины $\alpha = h/l$, где h — ширина опорного пучка, а l — расстояние между предметом и голограммой. Если $h = 3$ см, а $l = 30$ см, то должно быть $\Delta x < \lambda l/h = 5$ мкм.

С нелазерными источниками (например, ртутной дуговой лампой) этим требованиям пытались удовлетворить стандартными приемами, пропуская свет через различные монохроматоры, а затем фокусируя его на малом отверстии. При этом через отверстие проходила и достигала освещаемого объекта лишь ничтожная часть светового потока. Требующееся время экспозиции при этом будет достигать порядка года. Поэтому Габору удалось получить голограммы только простейших микроскопических объектов при экспозициях в несколько часов. Лазеры сделали описанные операции ненужными, а идея голографии получила практическое осуществление.

4. Представим поле волны, рассеянной предметом A, в виде

$$u = a(\mathbf{r})e^{i[\omega t - \Phi(\mathbf{r})]}, \tag{54.1}$$

а поле волны, отраженной плоским зеркалом S, — в виде

$$v = be^{i(\omega t - \mathbf{kr})} \tag{54.2}$$

(b и $\mathbf{k}$ — постоянные). Оба поля записаны в скалярной форме. Это упрощение не существенно, так как изменения поляризации волн в интересующих нас процессах не играют заметной роли. При таком упрощении величины $a(\mathbf{r})$, $\Phi(\mathbf{r})$ и b могут считаться вещественными. Интенсивность результирующего поля на пластинке Γ и перед ней представится выражением

$$I = (u + v)(u^* + v^*) = v^*u + vu^* + v^*v + u^*u.$$

Начало координат поместим в какой-либо точке O в плоскости пластинки Γ. С той же плоскостью совместим координатную плоскость XY, направив ось Y перпендикулярно к направлению опорного пучка, отраженного зеркалом S ($k_y = 0$). Тогда ось Z будет нормальна к плоскости пластинки Γ. Направим ее в сторону распространения опорного пучка. Интенсивность I_0 светового поля в плоскости пластинки найдется из предыдущего выражения, если все входящие в него функции заменить их значениями при $z = 0$:

$$I_0 = ba(x, y, 0)e^{i[k_x x - \Phi(x,y,0)]} + ba(x, y, 0)e^{-i[k_x x - \Phi(x,y,0)]} + \\ + b^2 + a^2(x, y, 0). \tag{54.3}$$

Допустим теперь, что контактным способом изготовлен позитив голограммы при таком режиме проявления, что амплитудная прозрачность (пропускаемость) D позитива оказалась пропорциональной I_0. Для краткости коэффициент пропорциональности можно принять равным единице, т.е. положить $D = I_0$. Этого всегда можно достигнуть надлежащим выбором единиц. Такая позитивная голограмма может быть использована для восстановления рассеянной волны $u(\mathbf{r}, t)$. Для этого, убрав предмет A, голограмму просвечивают таким же опорным пучком $v(\mathbf{r}, t)$, какой применялся при ее изготовлении. Этот пучок будет испытывать дифракцию на голограмме подобно тому, как дифрагирует свет на дифракционной решетке. Задача сводится к расчету дифракционной картины за голограммой. Для ее решения можно применить метод Рэлея. Обозначим через $E(\mathbf{r}, t)$ волновое поле за голограммой. На выходе голограммы оно представится выражением

$$E_{\text{вых}} = Dv(x, y, 0) = I_0 be^{i(\omega t - k_x x)}. \tag{54.4}$$

Надо найти решение волнового уравнения

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} + k^2 E = 0, \tag{54.5}$$

удовлетворяющее краевому условию (54.4). Ввиду линейности и однородности такого уравнения и условия (54.4), эту задачу можно расчленить на четыре независимые задачи. Представим функцию E в виде суммы

$$E = E_1 + E_2 + E_3 + E_4 \tag{54.6}$$

и потребуем, чтобы слагаемые E_1, E_2, E_3, E_4 были решениями уравнения (54.5) и удовлетворяли следующим граничным условиям:

$$E_{1\text{ вых}} = b^2 a(x, y, 0) e^{i[\omega t - \Phi(x,y,0)]}, \tag{54.7}$$

$$E_{2\text{ вых}} = b^2 a(x, y, 0) e^{i[\omega t + \Phi(x,y,0) - 2k_x x]}, \tag{54.8}$$

$$E_{3\text{ вых}} = b^3 e^{i(\omega t - k_x x)}, \tag{54.9}$$

$$E_{4\text{ вых}} = b a^2(x, y, 0) e^{i(\omega t - k_x x)}. \tag{54.10}$$

Сюда необходимо добавить физическое условие, чтобы волны, дифрагированные на голограмме, были *уходящими*. Этим обеспечивается *единственность решения*.

5. Проще всего найти функцию E_3, т.е. решение задачи, удовлетворяющее граничному условию (54.9). Ввиду постоянства b таким решением, очевидно, будет

$$E_3 = b^3 e^{i(\omega t - \mathbf{kr})} = b^2 v(\mathbf{r}, t). \tag{54.9а}$$

Если отвлечься от амплитудного множителя b^2, то это есть просвечивающая (опорная) волна, распространившаяся за голограмму.

6. Основной интерес в голографии представляет решение E_1, удовлетворяющее граничному условию (54.7). Ввиду постоянства b, его можно также сразу указать. Именно:

$$E_1 = b^2 a(x, y, z) e^{i[\omega t - \Phi(x,y,z)]} = b^2 u(\mathbf{r}, t). \tag{54.7а}$$

Действительно, функция E_1 удовлетворяет волновому уравнению (54.5) и представляет волну, уходящую от голограммы, которая при $z = 0$ переходит в выражение (54.7). Но эта волна с точностью до несущественного постоянного множителя тождественна с волной u, рассеянной объектом. Она, таким образом, точно воспроизводит рассеянную волну и дает *мнимое изображение* (МИ) объекта, в том же самом месте, в каком он находился во время получения голограммы (рис. 207 *а*). Тем самым разъяснен основной принцип голографни.

7. Для нахождения волны E_2, т.е. решения волнового уравнения, удовлетворяющего граничному условию (54.8), рассмотрим сначала частный случай, когда опорный, а следовательно, и просвечивающий пучки перпендикулярны к плоскости голограммы. Тогда $k_x = 0$, и граничное условие (54.8) переходит в

$$E_{2\text{ вых}} = b^2 a(x, y, 0) e^{i[\omega t + \Phi(x,y,0)]}. \tag{54.11}$$

Волна

$$\widetilde{E}_2(x, y, z) = b^2 a(x, y, z) e^{i[\omega t + \Phi(x,y,z)]}, \qquad (54.12)$$

очевидно, удовлетворяет этому граничному условию. Это есть волна, получающаяся из волны (54.7а) обращением направления ее распространения, а потому она также удовлетворяет и волновому уравнению (54.5). Она давала бы мнимое изображение предмета перед голограммой в точности в том же месте, где

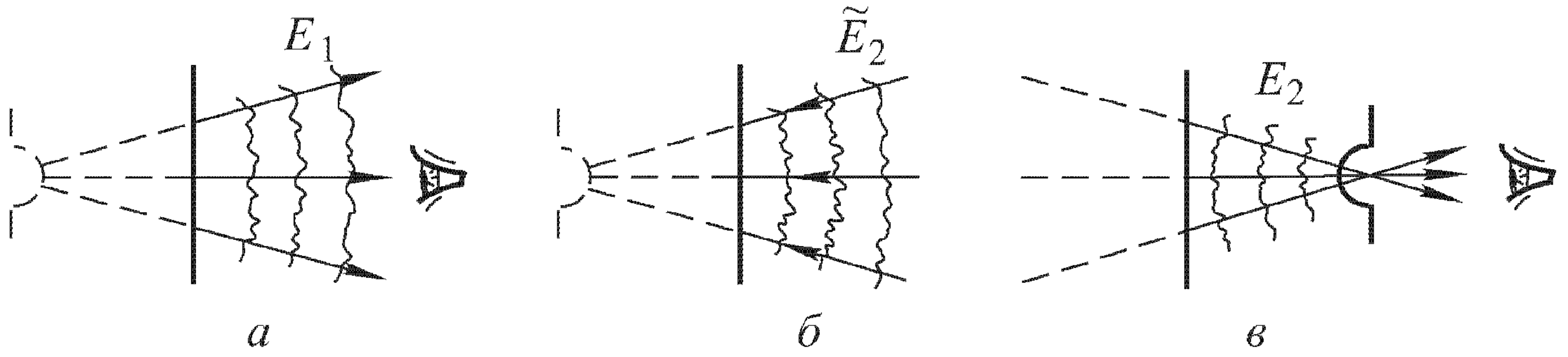


Рис. 207

находился сам предмет при получении голограммы (рис. 207 *б*). Однако волна $\widetilde{E}_2(\mathbf{r}, t)$ распространяется к голограмме, а не от нее, и по этой причине не может служить решением рассматриваемой нами задачи. Для нахождения нужного решения заметим, что волновое уравнение (54.5) не изменится, если изменить на противоположные знаки одной или нескольких координат. Так, при изменении знака у z уравнение (54.5) переходит в

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial (-z)^2} + k^2 E = 0. \qquad (54.5\text{а})$$

Это уравнение, очевидно, имеет решение $\widetilde{E}_2(x, y, -z)$. А так как оно тождественно с уравнением (54.5), то последнее имеет такое же решение, т.е.

$$E_2 = b^2 a(x, y, -z) e^{i[\omega t + \Phi(x,y,-z)]}. \qquad (54.11\text{а})$$

Решение это, очевидно, удовлетворяет граничному условию (54.11). Оно получится из (54.12), если пространства перед и за голограммой поменять местами. Решение (54.11а) представляет волну, *идущую от голограммы*, а потому оно и будет искомым решением задачи. Полученная волна сходится за голограммой и дает там *действительное изображение* (ДИ) предмета. Положение его зеркально симметрично с положением самого предмета относительно плоскости голограммы (рис. 207 *в*). Однако глаз может видеть действительное изображение только тогда, когда оно рассматривается с задней стороны, т.е. когда изображение находится между голограммой и глазом. Поэтому действительное изображение будет *псевдоскопическим*, или вывернутым. На нем выпуклые места предмета глаз видит как вогнутые и наоборот. (Псевдоскопичность не будет проявляться, если действительное изображение спроецировать на экран.) Если смотреть на голограмму перпендикулярно к ее поверхности, то действительное изображение будет мешать рассматриванию мнимого изображения, так как оба изображения лежат на одной линии зрения. Такое же вредное влияние оказывают прямые лучи просвечивающего пучка, поскольку они также попадают в глаз. Эти недостатки устраняются применением наклонных опорных (а следовательно, и просвечивающих) пучков света.

В случае наклонных пучков $k_x \neq 0$, и надо найти решение E_2, удовлетворяющее общему граничному условию (54.8). Но этот случай легко сводится к частному случаю нормального падения, разобранному выше. Действительно, если бы граничное условие имело частный вид (54.11), то можно было бы перейти к общему граничному условию (54.8), наложив на голограмму дополнительную дифракционную решетку с амплитудной прозрачностью $D = e^{-2ik_x x}$. Это — фазовая решетка, дающая только спектр минус первого порядка (см. § 52, п. 5). Она действует как призма, т.е. только изменяет направление падающего пучка света. В результате действительное изображение смещается в сторону, как указано на рис. 208 *а*. Такое смещение, конечно, искажает изображение совершенно так же, как это происходит при рассматривании предмета через призму. При большом наклоне опорного (а следовательно, и просвечивающего) пучка действительного изображения может и не получиться.

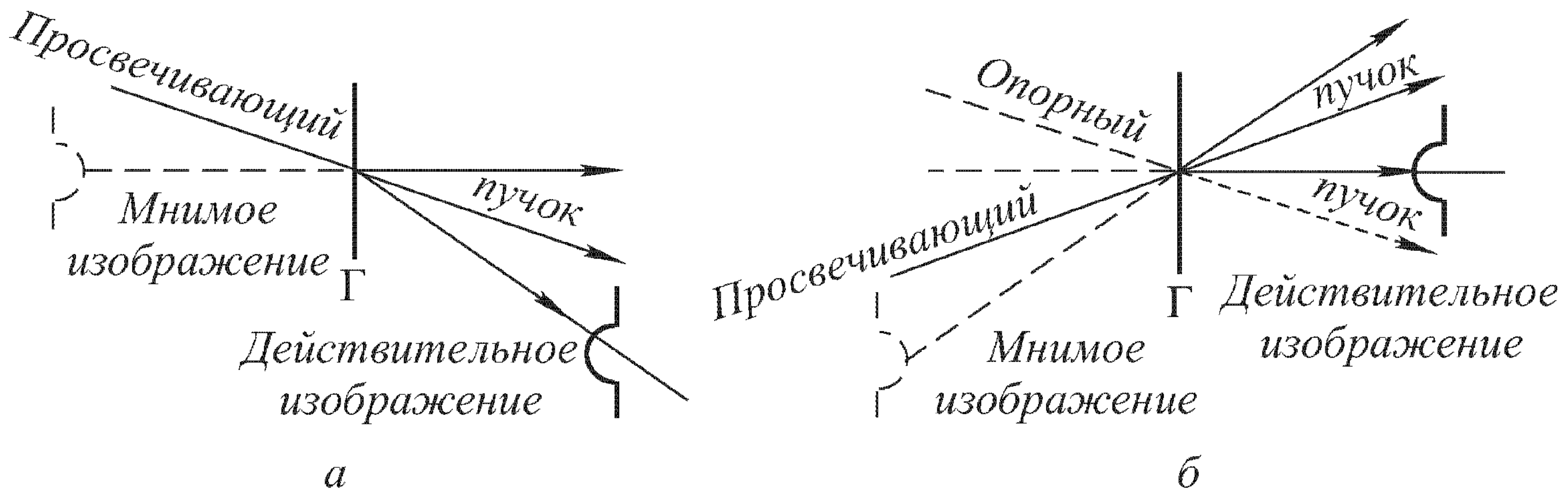


Рис. 208

8. Искажения действительного изображения можно избежать, если просвечивающий пучок направить на голограмму под тем же углом и в той же плоскости падения, что и опорный, но с другой стороны от нормали к плоскости голограммы. Тогда

$$v_{\text{прос}} = b e^{i(\omega t + k_x x - k_z z)},$$

и вместо граничных условий (54.7) и (54.8) мы получим соответственно

$$E_{1\text{ вых}} = b^2 a(x, y, 0) e^{i[\omega t - \Phi(x,y,0) + 2k_x x]}, \tag{54.7б}$$

$$E_{2\text{ вых}} = b^2 a(x, y, 0) e^{i[\omega t + \Phi(x,y,0)]}. \tag{54.8б}$$

Теперь действительное изображение будет расположено зеркально симметрично (относительно плоскости голограммы) с самим предметом в том положении, какое он занимал при изготовлении голограммы. Оно по-прежнему будет псевдоскопическим, но не искаженным. Зато мнимое изображение сместится в сторону (рис. 208 *б*) и будет искажено, как при рассматривании предмета через призму.

9. Остается выяснить влияние волны E_4, удовлетворяющей граничному условию (54.10). Найти решение в этом случае без конкретизации вида функции $a(x, y, 0)$ невозможно, так как в условие (54.10) входит не сама эта функция, а ее квадрат. Для отыскания E_4 в общем случае следует разложить функцию $a^2(x, y, 0)$ в интеграл Фурье, а затем стандартным способом (см. § 52) представить волну E_4 в виде суперпозиции плоских волн, распространяющихся в различных направлениях. Таким образом, за голограммой возникает волновое поле, создающее фон, на котором получаются мнимое и действительное изображения. Интенсивность этого фона зависит от квадрата амплитуды $a(x, y, z)$ волны, рассеянной предметом при изготовлении голограммы. Поэтому для ослабления вредного влияния такого фона интенсивность опорной волны берут значительно большей интенсивности предметной волны.

Так как фотослой, на котором записана голограмма, тонкий, то при восстановлении изображения не возникает интерференция столь же высоких порядков, что и при изготовлении голограммы. По этой причине удовлетворительное изображение предмета можно получить с голограммой, освещаемой при восстановлении светом от ртутной лампы.

10. Чем больше углы, под которыми сходятся интерферирующие лучи на голограмме, а следовательно, чем мельче детали интерференционной картины, тем выше разрешающая способность голограммы (см. § 57). При предельно возможном разрешении эти детали порядка длины световой волны. Но и при углах схождения в несколько градусов интерференционная картина на голограмме получается все же довольно тонкой. Чтобы ее зафиксировать на фотопластинке, требуется фотографическая эмульсия весьма высокого качества. Наиболее высокие требования к фотопластинкам предъявляются в цветной фотографии Денисюка, где предметная и опорная волны идут навстречу друг к другу (см. п. 13). Современные мелкозернистые фотопластинки, применяемые в голографии, разрешают 1000–

10 000 линий на миллиметр (расстояние между разрешаемыми линиями 1000–100 нм). Следует, однако, заметить, что повышение разрешающей способности фотоэмульсии, как правило, сопровождается падением ее чувствительности.

11. Для получения изображения нет необходимости изготовлять позитивную голограмму; негатив столь же хорош, что и позитив. Это непосредственно следует из теоремы Бабине. Различие в дифракционных картинах касается только просвечивающего пучка, проникшего за голограмму, и совсем не затрагивает волновые поля E_1 и E_2, дающие мнимое и действительное изображения. В обоих случаях при восстановлении волнового фронта получаются одинаковые изображения.

Конечно, при изготовлении голограммы условие $D \sim I_0$, соблюдение которого предполагалось при нашем изложении, не может быть выполнено вполне точно. Его невыполнение приводит к появлению *дополнительных изображений*. Это проще всего пояснить на примере, в котором предметная и опорная волны обе плоские и падают на фотопластинку под разными углами (предметом является бесконечно удаленная светящаяся точка). Тогда при соблюдении условия $D \sim I_0$ голограмма будет представлять собой дифракционную решетку с синусоидальной амплитудной прозрачностью. При ее просвечивании, наряду со спектром нулевого порядка (т.е. прямого просвечивающего пучка, проникшего за голограмму), появятся спектры первого и минус первого порядков, один из которых будет мнимым, а другой действительным изображениями предмета. Если же условие $D \sim I_0$ не соблюдается, то разложение функции $D(x)$ в ряд Фурье будет содержать гармоники высших порядков. При просвечивании голограммы, наряду со спектрами первых порядков, появятся спектры высших порядков, т.е. дополнительные мнимые и действительные изображения предмета. Однако если такие дополнительные изображения слабы и получаются в стороне от основного изображения, то они не причиняют существенного вреда.

Для получения большей яркости изображения применяют так называемые *отбеленные голограммы*. Они получаются из обычных голограмм путем обработки их специальными отбеливающими растворами, образующими прозрачные соединения с серебром, отложившимся в эмульсии голограммы. Получается прозрачная фазовая голограмма, пропускаемость которой варьируется как за счет толщины, так и за счет показателя преломления отбеленного слоя. Такая голограмма сохраняет способность давать оптические изображения. В какой степени отбелка голограммы и отступления от условия $D \sim I_0$ сказываются на получаемых изображениях, эти вопросы представляют громадный интерес в практической голографии. Однако они трудно поддаются теоретическому изучению и по своему содержанию не могут быть включены в курс общей физики.

12. По сравнению с обычными фотографиями изображения, получаемые по методу голографии, обладают тем преимуществом, что они *трехмерны*. Мнимые изображения полностью воспроизводят без каких бы то ни было искажений взаимное расположение реальных предметов в пространстве. Если при рассматривании голограммы одни предметы заслоняются другими, то достаточно сместить в сторону глаз, чтобы увидеть и заслоненные предметы. Часть голограммы действует как целая голограмма. Например, для воспроизведения изображения годится каждый кусочек разбитой голограммы. По мере уменьшения размеров голограммы ослабевают лишь четкость изображения (разрешающая способность) и ощущение объемности. Эта особенность голографического метода связана с тем, что при экспонировании, как правило, все части фотопластинки подвергаются действию света, рассеянного *всеми точками предмета*. Поэтому в каждой части голограммы хранится в закодированной форме изображение всего предмета. Вопрос о разрешающей способности голограмм будет разобран в § 57 (п. 5), как частный случай общего вопроса о разрешающей способности оптических приборов.

На одной фотопластинке можно последовательно зарегистрировать несколько изображений, причем каждое из них можно восстановить без помех со стороны других изображений. Можно, например, произвести несколько экспозиций при одном и том же опорном пучке, голографируя каждый раз только один предмет. Во избежание наложения восстановленных изображений друг на друга голографируемые объекты должны располагаться в различных местах. Можно также объекты помещать в одно и то же место, но менять направление опорного пучка. Можно получать цветные изображения, освещая предмет тремя источниками света с различными длинами волн, подобранными так, чтобы наиболее полно передать цвет предмета. От тех же трех источников формируются опорные пучки, создающие вместе с волнами, рассеянными предметом, единую голограмму на черно-белой фотопластинке. Для восстановления изображения такая «трехкомпонентная голограмма» помещается в то же место, какое она занимала при экспонировании, и одновременно освещается тремя просвечивающими пучками, тождественными с опорными. При просвечивании голограммы каждый из этих пучков создает три действительных и три мнимых изображения. Три из этих мнимых изображений, и притом разноцветных, пространственно совпадут друг с другом, образуя цветное изображение. Остальные изображения расположатся в разных местах, не мешая основному изображению.

13. Голограммы, о которых шла речь выше, можно назвать *двумерными*, так как в них применяются фотопластинки с тонкими слоями эмульсии. В 1962 г. советский физик Ю.Н. Денисюк (р. 1927) стал получать *трехмерные*, *объемные* голограммы на фотопластинках с *толстослойной эмульсией*. В его методе удачно сочетаются принципы голографии и цветной фотографии Липпмана. Толщина фотографического слоя составляет 15–20 мкм, т.е. 30–40 длин волн зеленого цвета. Фотослой настолько прозрачен, что через него можно освещать голографируемый объект. Опорная плоская монохроматическая волна от лазера падает на фотопластинку со стороны стекла (рис. 209). Пройдя через фотопластинку, она освещает голографируемый предмет. Волна, рассеянная предметом, распространяется навстречу опорной волне, интерферируя с ней в толще фотоэмульсии. Интерференционная картина представляет *стоячие волны*, на которые наложен довольно причудливый узор мелких деталей из максимумов и минимумов, так как среди интерферирующих волн только опорная волна является плоской. Проявленная и отфиксированная фотопластинка и будет объемной голограммой Денисюка. Она состоит как бы из нескольких десятков поверхностных голограмм, расположенных в толще эмульсии.

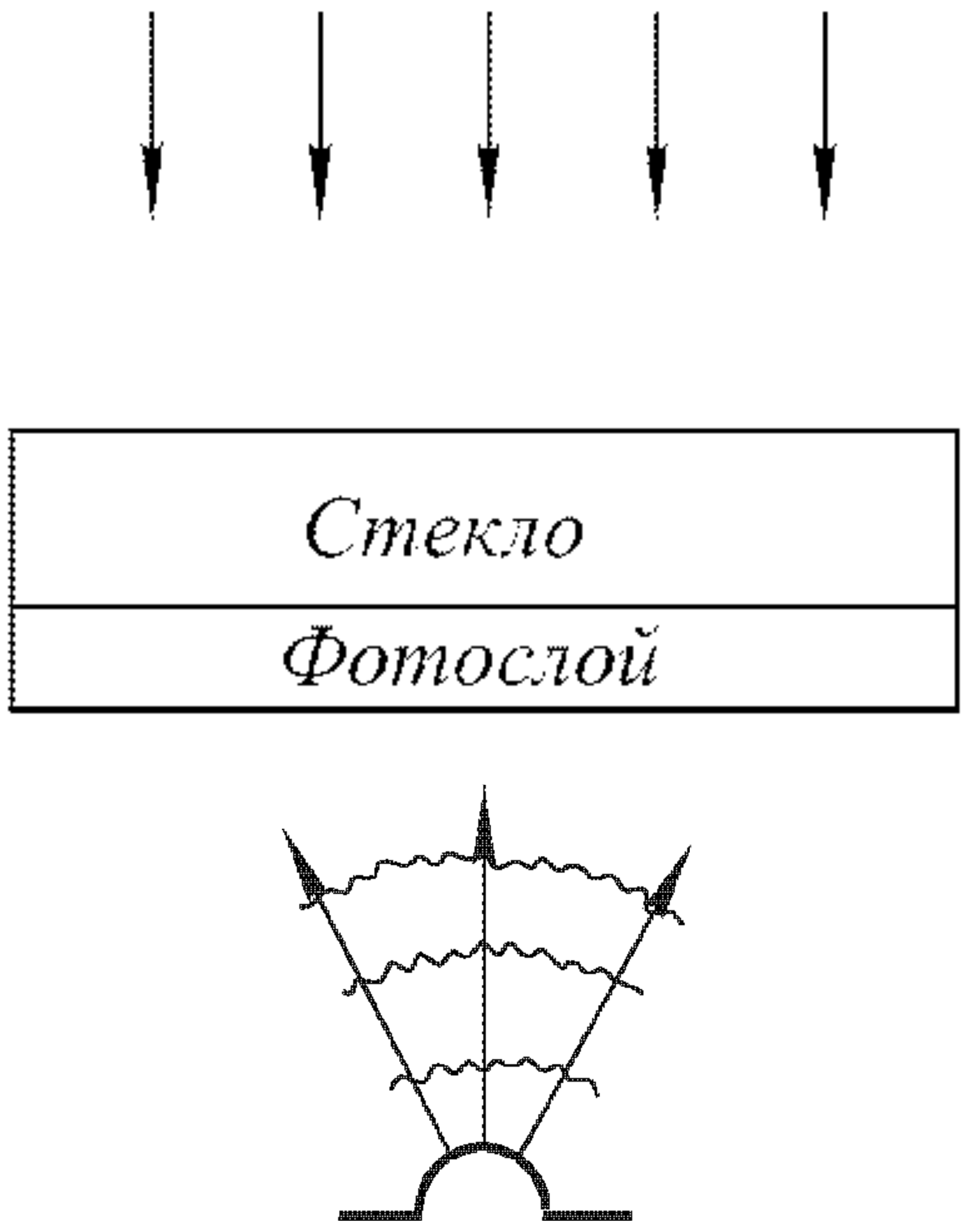


Рис. 209

Восстановление предметной волны производится расходящимся пучком *белого света*. Каждый слой выделившегося серебра, действуя подобно двумерной голограмме, дает слабые мнимое и действительное изображения предмета. При многолучевой интерференции происходит усиление тех волн, длина которых равна длине волны излучения лазера, в тех направлениях, в которых разность фаз между волнами от соседних слоев серебра равна 2π. В результате возникают изображения того же цвета, что и цвет луча лазера. Остальные изображения гасят друг друга при интерференции.

Таким образом, голограмма производит *монохроматизацию* белого света, которым она освещается. Конечно, такая монохроматизация сравнительно невысокая, из-за незначительного числа отложившихся слоев серебра и связанной с этим небольшой спектральной разрешающей способности голограммы. Кроме того, цвет изображения может существенно отличаться от цвета излучения лазера. Это связано с изменением расстояний между слоями почернения при проявлении, фиксировании и сушке фотопластинки.

Метод Денисюка, подобно трехцветной фотографии, позволяет получать изображения предметов в *натуральных цветах.* Для этого на одной и той же фотопластинке получают голограмму предмета с помощью трех лазеров, излучения которых имеют различные длины волн. Последние подбираются так, чтобы при смешении они наиболее совершенно воспроизводили цвет предмета. Такая голограмма действует как три голограммы, дающие при освещении белым светом совмещенные изображения предмета в трех цветах. При этом цвет изображения кажется глазу таким же, как и цвет самого предмета.

Голография в настоящее время представляет самостоятельный, быстро развивающийся раздел науки, техники и искусства, возможно с большим будущим. Ей посвящены специальные руководства, к которым и следует обратиться всем, кто пожелает глубже и детальнее изучить голографию.

§ 55. Световое поле вблизи фокуса

1. Согласно геометрической оптике, волновое поле в центре сферической сходящейся волны (фокусе) обращается в бесконечность. Это указывает на неприменимость геометрической оптики в фокусе и его ближайшей окрестности.

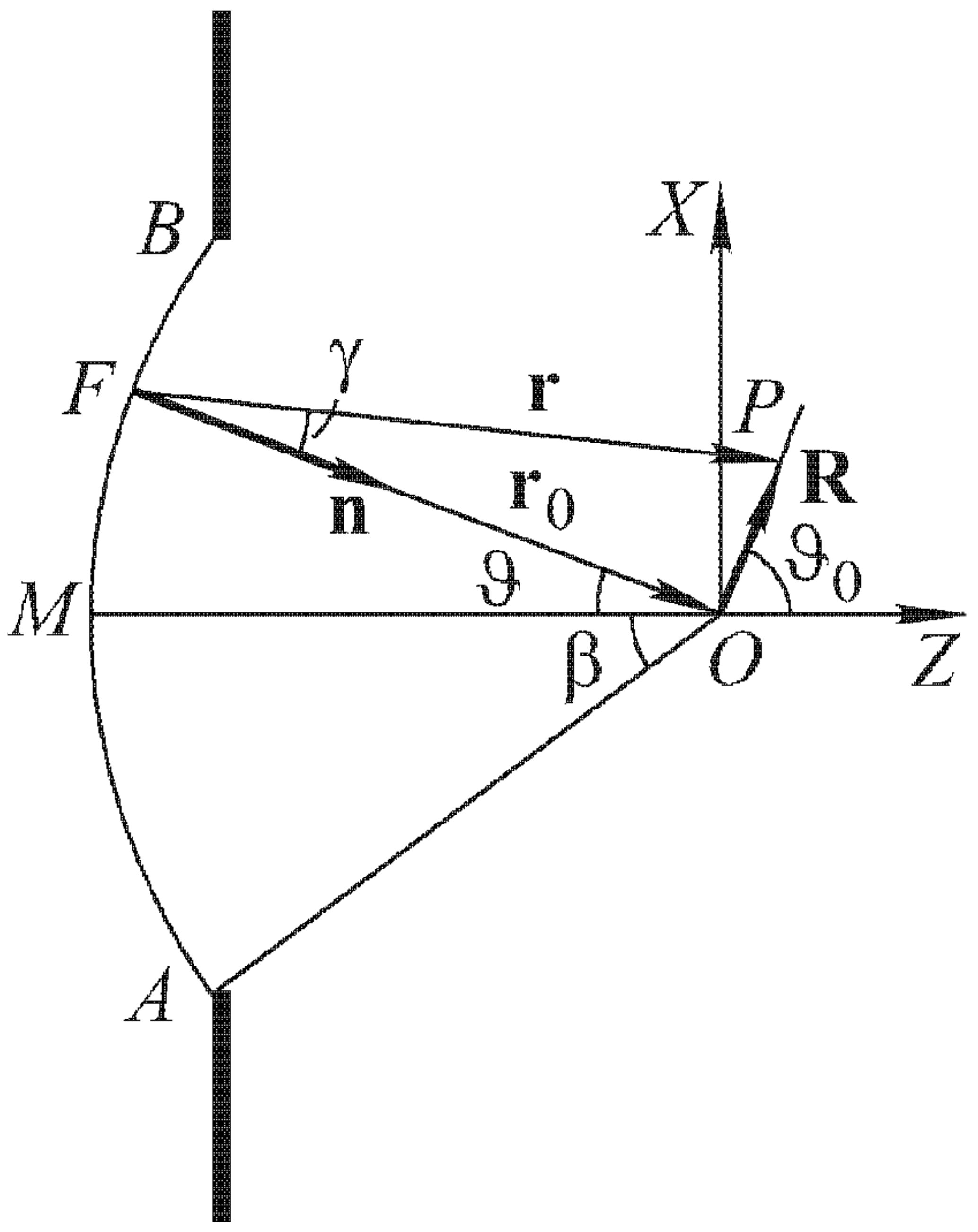


Рис. 210

Рассмотрим задачу о волновом поле в окрестности фокуса с точки зрения волновой оптики. Пусть на пути сходящейся сферической волны поставлена диафрагма с отверстием AB (рис. 210). Неприкрытую часть волнового фронта F примем за вспомогательную поверхность, из которой исходят вторичные волны Гюйгенса. Следуя приближенному методу Френеля (§ 41), поле на поверхности F запишем в виде

$$E_F = \frac{1}{r_0} e^{i(\omega t + kr_0)}.$$

Если пренебречь зависимостью амплитуд вторичных волн от (малого) угла γ, то поле в точке наблюдения P представится интегралом

$$E_P \sim e^{i\omega t} \int \frac{e^{ik(r_0 - r)}}{r_0 r}\, dF,$$

распространенным по всей неприкрытой поверхности волнового фронта. Если точка P лежит в малой окрестности фокуса O, то расстояния r_0 и r в знаменателе подынтегрального выражения могут считаться постоянными. Различие между ними надо учесть только в фазовом множителе $e^{ik(r_0-r)}$. Как видно из рис. 210, $\mathbf{r} = \mathbf{r}_0 + \mathbf{R}$. Возведем это равенство в квадрат, учтем, что $R \ll r_0$, а затем приближенно извлечем квадратный корень с точностью до квадратичных членов по R включительно. Получим

$$r - r_0 = \frac{(\mathbf{r}_0\mathbf{R})}{r_0} + \frac{R^2}{2r_0} - \frac{(\mathbf{r}_0\mathbf{R})^2}{r_0^3} = (\mathbf{nR}) + \frac{R^2}{2r_0}\sin^2(\mathbf{R}, \mathbf{n}),$$

где $\mathbf{n} = \mathbf{r}_0/r_0$ — единичный вектор нормали к поверхности F, направленный к фокусу O. В дальнейшем в этом выражении сохраним только линейный член. Это можно делать, если поправка в фазе, вносимая квадратичным членом, много меньше π, т.е. когда

$$R^2/r_0 \ll \lambda. \tag{55.1}$$

С учетом всего изложенного для волнового поля (в условных единицах) в точке P получим

$$E_P = \int e^{i(\omega t - k\mathbf{nR})}\, d\Omega, \tag{55.2}$$

где $d\Omega = dF/r_0^2$ — телесный угол, под которым из точки O видна площадка dF.

Когда диафрагма удалена в бесконечность, а расстояние R остается конечным, то $R^2/r_0 = 0$. Тем самым ограничение (55.1) снимается, а (55.2) переходит в точное решение задачи. Это видно и из самой структуры выражения (55.2). Действительно, $d\Omega\, e^{i(\omega t - k\mathbf{nR})}$ есть плоская волна, распространяющаяся в направлении волнового вектора $\mathbf{k} = k\mathbf{n}$, а потому она удовлетворяет волновому уравнению (43.1). Интеграл (55.2) есть суперпозиция таких плоских волн различных направлений, а следовательно, является точным решением того же уравнения. Такое решение позволяет точно исследовать все явления, связанные с прохождением сферической волны через фокус.

2. Если точка P находится в фокусе O, т.е. $R = 0$, то все вторичные волны приходят в O в одинаковых фазах, а потому интенсивность света в этой точке максимальна. Формула (55.2) в этом случае дает $E_O = \Omega e^{i\omega t}$. Наибольший интерес представляет распределение интенсивности света в фокальной плоскости, т.е. в плоскости, проходящей через фокус O перпендикулярно к оси пучка MO. Чтобы интегрирование в (55.2) выполнялось элементарно, будем предполагать телесный угол Ω небольшим, так что при вычислении интеграла неприкрытую часть сферического волнового фронта можно заменить плоской. Саму диафрагму возьмем прямоугольной формы со сторонами $2a$ и $2b$. Угловые размеры этих сторон при наблюдении из точки O будут $2\alpha = 2a/r_0$ и $2\beta = 2b/r_0$ соответственно. Начало координат поместим в точке O, направив оси X и Y параллельно сторонам прямоугольной диафрагмы. Координаты точки наблюдения P обозначим через X и Y, текущие координаты в плоскости диафрагмы — через x и y. Тогда $dF = dx\,dy = r_0^2\,d\varphi\,d\psi$, $d\Omega = d\varphi\,d\psi$, где φ и ψ — углы, под которыми из точки O видны отрезки x и y. Далее, $(\mathbf{nR}) = n_x X + n_y Y = -(x/r_0)X - (y/r_0)Y = -(\varphi X + \psi Y)$. В результате интеграл (55.2) перейдет в

$$E_P = e^{i\omega t} \int\limits_{-\alpha}^{+\alpha} \int\limits_{-\beta}^{+\beta} e^{ik(X\varphi + Y\psi)}\,d\varphi\,d\psi.$$

Это в точности такой же интеграл, какой встречался в § 45 при рассмотрении фраунгоферовой дифракции на прямоугольном отверстии. Поэтому распределение интенсивности света в фокальной плоскости в окрестности фокуса можно представить формулой

$$I = I_0 \left(\frac{\sin kx\alpha}{kx\alpha}\right)^2 \left(\frac{\sin ky\beta}{ky\beta}\right)^2. \qquad (55.3)$$

В фокальной плоскости получается система светлых пятен прямоугольной формы со светлым центром (см. рис. 179). Расстояния между двумя соседними минимумами, а также от центра центрального максимума до первого минимума равны

$$\Delta x = \frac{\pi}{k\beta} = \frac{1}{2}\frac{\lambda}{\beta}. \qquad (55.4)$$

Остается проверить, выполняется ли условие (55.1). Так как заметная интенсивность по формуле (55.3) получается при R порядка Δx, то формула (55.1) переходит в

$$\beta \gg \sqrt{\lambda/r_0}. \qquad (55.5)$$

В оптических приборах под r_0 следует понимать расстояние от линзы (или сферического зеркала) до точки геометрического схождения лучей. Например, если лучи сходятся в главном фокусе, то r_0 равно фокусному расстоянию f линзы или зеркала. Ввиду малости длины волны, условие (55.5) очень хорошо выполняется во всех оптических приборах. Так, при $f = 10$ см, $\lambda = 500$ нм из (55.5) получаем $\beta > 4 \cdot 10^{-3}$ рад $\sim 15'$. Поэтому применимость выведенных здесь формул к оптическим приборам с линзами и зеркалами не вызывает сомнений.

Для теории оптических приборов наибольшее значение имеет случай круглой диафрагмы. Исследование этого случая, конечно, не встречает каких-либо затруднений. Качественно ясно, что дифракционная картина в фокальной плоскости должна иметь вид светлых и темных концентрических кругов со светлым центром (см. рис. 180). Для определения размеров дифракционных кругов надо вычислить интеграл (55.2). В случае круглой диафрагмы результат вычисления выражается через бесселеву функцию первого порядка. Радиусы темных колец имеют следующие значения:

$$R = 0{,}61\lambda/\beta; \quad 1{,}12\lambda/\beta; \quad 1{,}62\lambda/\beta; \ldots, \tag{55.6}$$

где β — угловой размер радиуса диафрагмы, если его рассматривать из точки O.

3. Найдем теперь распределение интенсивности света на оси MO, предполагая, что пучок лучей ограничен круглой диафрагмой. В этом случае в (55.2) можно взять $d\Omega = 2\pi \sin\vartheta\, d\vartheta$ (см. рис. 210). Если точка P лежит перед фокусом, то $\mathbf{nR} = -R\cos\vartheta$. Если же она лежит за фокусом, то $\mathbf{nR} = R\cos\vartheta$. По формуле (55.2)

$$E_P = 2\pi e^{i\omega t} \int\limits_0^\beta e^{\pm ikR\cos\vartheta} \sin\vartheta\, d\vartheta = \frac{2\pi e^{i(\omega t \pm kR)}}{\pm ikR}\left[1 - e^{\pm ikR(\cos\beta - 1)}\right],$$

где верхний знак относится к случаю, когда точка наблюдения расположена левее, а нижний — правее фокуса. Вычислив модуль этого выражения, находим амплитуду колебаний поля на оси пучка:

$$a \sim \frac{1}{R}\sin\frac{kR(1-\cos\beta)}{2}. \tag{55.7}$$

Во всякой плоскости, перпендикулярной к оси пучка, дифракционная картина имеет вид концентрических колец. Однако в зависимости от положения этой плоскости центр пучка может быть и светлым, и темным. Амплитуда на оси обращается в нуль, когда $kR(1-\cos\beta) = m\pi$, т.е.

$$R = \frac{m\lambda}{1-\cos\beta} \qquad (m = 1, 2, 3, \ldots). \tag{55.8}$$

В этом случае, как легко убедиться, в неприкрытой части сферического волнового фронта содержится четное число зон Френеля, а потому центр колец и получается темным. Расстояние между ближайшими минимумами интенсивности, расположенными по разные стороны от фокуса, равно

$$\delta = \frac{2\lambda}{1-\cos\beta} \approx \frac{4\lambda}{\beta^2}. \tag{55.9}$$

Эта величина может служить мерой продольных размеров области, в которой концентрируется свет вблизи фокуса. Объем этой области порядка $\frac{4\lambda}{\beta^2}\cdot\left(\frac{\lambda}{\beta}\right)^2 = \frac{4\lambda^3}{\beta^4}$.

§ 56. Разрешающая способность телескопа и микроскопа

1. Будем предполагать, что оптические системы идеальны в смысле геометрической оптики, т.е. дают строго точечные изображения каждой точки объекта. По волновой оптике это не так. Как показано в предыдущем параграфе, изображением светящейся точки в сопряженной плоскости является дифракционная картина, состоящая из концентрических колец, окружающих центральный светлый дифракционный кружок. Распределение интенсивности в такой картине представлено сплошной кривой на рис. 181. Основная доля энергии света (около 84 %) приходится на центральный дифракционный кружок. Этот кружок и будет изображением светящейся точки, если пренебречь энергией, приходящейся на окружающие его дифракционные кольца. Он называется *кружком Эйри* по имени ученого, впервые рассчитавшего дифракционную картину Фраунгофера от круглого отверстия.

Конечный объект можно рассматривать как совокупность точечных источников, каждый из которых изображается кружком Эйри (1801–1892) с окружающими его дифракционными кольцами. Изображение объекта есть наложение таких кружков и дифракционных колец. Задача теории сводится к расчету распределения интенсивности света в такой картине. Следует различать два предельных случая: 1) *точечные источники некогерентны*; 2) *точечные источники когерентны*. В первом случае складываются *интенсивности* волновых полей, во втором — их *напряженности*. Приближенно первый случай реализуется для *самосветящихся*, второй — для освещаемых объектов. Первый случай имеет основное значение в теории телескопа, а второй — микроскопа.

В простейшем случае сложный объект состоит из двух точечных источников S_1 и S_2 (рис. 211). Если расстояние между центрами кружков Эйри от этих точечных источников мало по сравнению с размерами самих кружков, то распределение интенсивности света в изображении будет мало отличаться от того, которое получилось бы только от одного точечного источника. Глаз не сможет решить, является ли объект точечным или состоит из нескольких близко расположенных светящихся точек. Говорят, что в этом случае светящиеся точки *не разрешаются* оптическим прибором. При увеличении расстояния между S_1 и S_2 будет увеличиваться и рас-

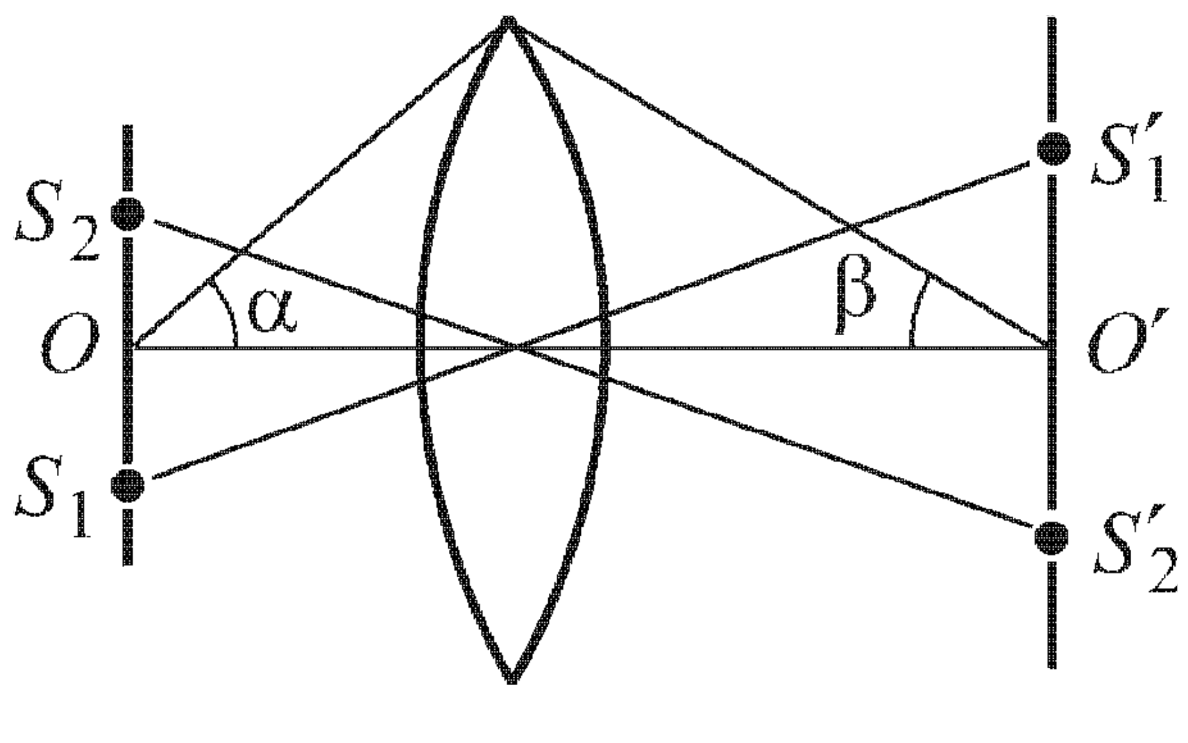


Рис. 211

стояние между центрами соответствующих им кружков Эйри S_1' и S_2', тогда как размеры самих кружков останутся неизменными. Начиная с некоторого минимального расстояния $l = l_{\text{мин}}$ на кривой распределения интенсивности в ее центре появится провал, который сможет зарегистрировать глаз или другой приемник света. Тогда говорят, что светящиеся точки S_1 и S_2 *разрешаются* оптическим прибором. Величина $l_{\text{мин}}$ называется *разрешаемым расстоянием прибора* (объектива), а обратная ему величина $1/l_{\text{мин}}$ — *разрешающей способностью.*

Точно указать величину $l_{\text{мин}}$ невозможно. Она в значительной степени зависит от индивидуальных особенностей глаза или другого приемника излучения. Речь может идти только о рациональной оценке $l_{\text{мин}}$. Согласно Рэлею, за $l_{\text{мин}}$ принимается такое расстояние между светящимися точками S_1 и S_2, при котором расстояние между центрами кружков Эйри равно радиусу одного кружка. В этом случае центр дифракционной картины от одной светящейся точки будет накладываться на первый дифракционный минимум дифракционной картины от второй светящейся точки. Такой критерий аналогичен критерию спектрального разрешения, предложенному также Рэлеем (см. § 47).

2. Для исследования вопроса о применимости критерия Рэлея к различным условиям наблюдения надо рассчитать распределение интенсивности света в плоскости изображения двух точечных источников света S_1 и S_2. С этой целью можно воспользоваться формулами предыдущего параграфа. Если источники S_1 и S_2 некогерентны, то надо сложить интенсивности создаваемых ими дифракционных картин. Если же они когерентны, то надо складывать волновые поля, а затем вычислять интенсивность результирующего волнового поля. Распределение интенсивности света в плоскости изображения в случае одной светящейся точки представлено сплошной кривой на рис. 181. Для нахождения распределения интенсивности света в случае двух одинаковых светящихся точек S_1 и S_2 применим графический метод.

Рассмотрим сначала случай, когда источники S_1 и S_2 некогерентны. На рис. 212 штриховые кривые представляют распределение интенсивности в плоскости изображения для каждого из точечных источников. Расстояние между центрами дифракционных картин предполагается равным $0{,}61\lambda/\beta$, т.е. радиусу кружка Эйри, как этого требует критерий Рэлея на пределе разрешения (угол β имеет тот же смысл, что и в предыдущем параграфе). Результирующая интенсивность получается сложением ординат обеих штриховых кривых и изображена сплошной линией. В центре C она имеет провал, где интенсивность составляет 74 % от максимальной интенсивности. Между тем средний глаз уверенно обнаруживает наличие провала уже тогда, когда интенсивность в центре составляет около 85 %. Таким образом, можно считать, что при выбранном расстоянии между светящимися точками получится разрешение. С другой стороны, сближая штриховые кривые, легко установить, что провал исчезает, когда расстояние между центрами кружков Эйри уменьшается до λ/β, что очень мало отличается от рэлеевского предела $0{,}61\lambda/\beta$. В этом случае разрешения нет. Таким образом, в случае некогерентных точечных источников света при визуальном наблюдении критерий Рэлея дает практически необходимое и достаточное условие для их разрешения.

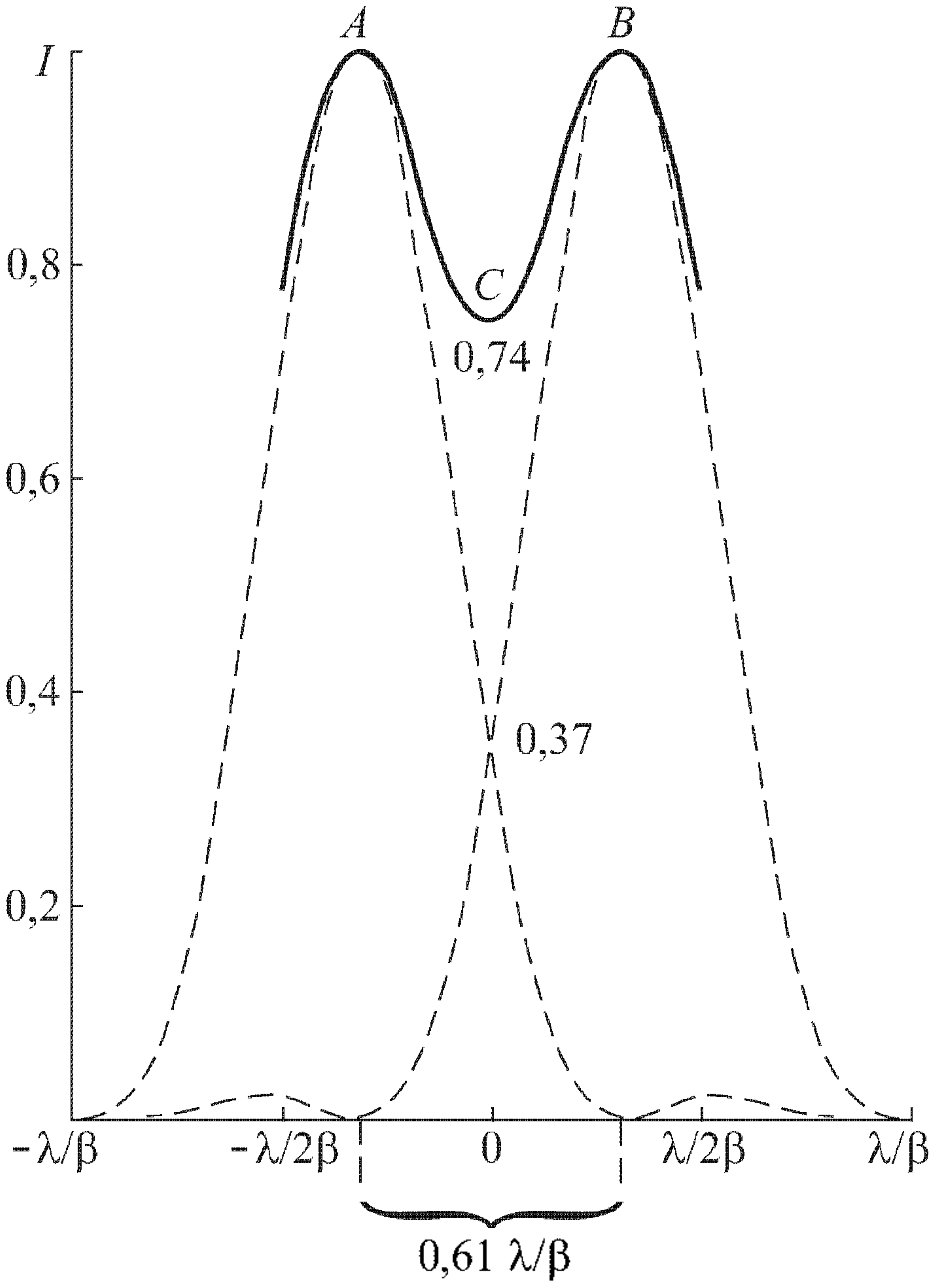


Рис. 212

3. Рассмотрим теперь случай, когда источники S_1 и S_2 когерентны. В качестве таких источников можно взять, например, два малых отверстия в непрозрачном экране, освещаемые когерентным светом (рис. 213). Такие отверстия ведут себя как точечные когерентные источники, излучающие вторичные волны Гюйгенса. Пусть отверстия освещаются пучком параллельных лучей, наклоненных под углом ϑ к нормали к плоскости экрана. Если l — расстояние между центрами отверстий, то разность фаз между волнами, приходящими в эти отверстия, будет $\Delta = (2\pi/\lambda)l\sin\vartheta$. Такова же будет и разность фаз между колебаниями, приходящими в центр дифракционной картины, т.е. в точку, расположенную посередине между центрами кружков Эйри от источников S_1 и S_2. Интенсивность света в этой точке представится выражением

$$I_C = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\Delta = 2I_1(1+\cos\Delta) = 4I_1\cos^2\frac{\Delta}{2}.$$

Принципиальное отличие от случая некогерентных источников света состоит в появлении дополнительного — интерференционного — члена. Это сказывается на критерии разрешения и на разрешающей способности объектива. Меняя угол наклона ϑ, можно повысить разрешающую способность объектива.

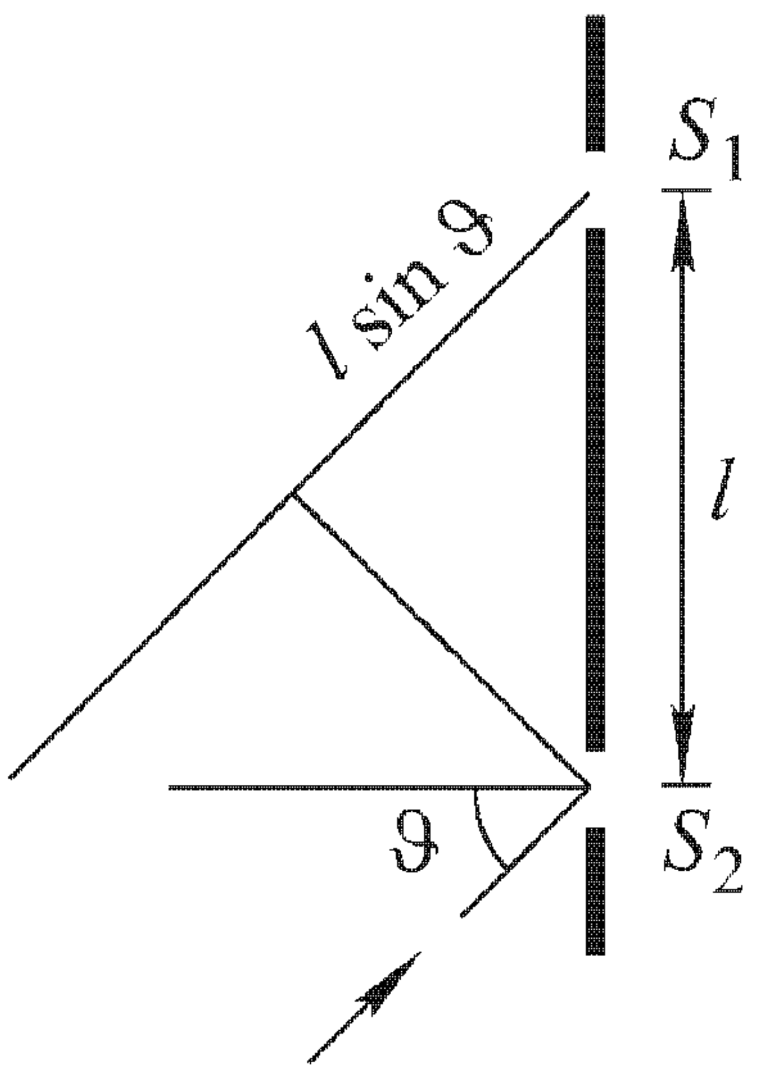


Рис. 213

Если свет, освещающий отверстия S_1 и S_2, падает нормально к плоскости экрана, то $\Delta = 0$ и, следовательно, $I_C = 4I_1$. Допустим, что в этом случае расстояние между центрами дифракционных кружков Эйри равно радиусу одного кружка, т.е. наименьшему расстоянию, которое требуется критерием Рэлея для разрешения светящихся точек. Тогда, как мы видели, интенсивность I_1 составляет 37 % от максимальной интенсивности, которую создает в центре кружка Эйри только один источник света. Поэтому результирующая интенсивность в центре дифракционной картины будет равна $4 \cdot 37 = 148\,\%$ от той же максимальной интенсивности. Распределение интенсивности света для рассматриваемого случая представлено на рис. 214. Штриховые кривые по-прежнему представляют интенсивности от каждого из источников S_1 и S_2 в отдельности, а сплошная кривая — результирующую интенсивность. На кривой результирующей интенсивности имеется всего лишь один максимум C. Поэтому никакого раздельного изображения светящихся точек в рассматриваемом случае получиться не может. Для этого необходимо развести центры кружков Эйри на расстояние, превышающее рэлеевский предел примерно в 1,2 раза.

Перейдем теперь к рассмотрению наклонного освещения. Если $\Delta = \pi/2$, т.е. разность хода между лучами, освещающими отверстия S_1 и S_2, равна $\lambda/4$, то $\cos\Delta = 0$. Интерференционный член пропадает, и получается в точности такое же распределение интенсивности света, а с ним и тот же критерий разрешения, что и в случае некогерентных светящихся точек. Если $\Delta = \pi$, а следовательно, $l = \lambda/(2\sin\vartheta)$, что возможно лишь при $l \geqslant \lambda/2$, то в центре дифракционной картины всегда получается нулевая интенсивность, а потому разделение максимумов выражено особенно отчетливо. На рис. 215 представлено распределение интенсивностей для рассматриваемого случая в предположении, что центры кружков Эйри разведены до рэлеевского предела. При скользящем падении лучей эти центры можно сблизить примерно вдвое по сравнению с рэлеевским пределом, и все же провал на кривой интенсивности сохранится. Таким образом, применение наклонного освещения приводит к повышению разрешающей способности. При скользящем падении разрешающая способность возрастает почти вдвое по сравнению со случаем некогерентных светящихся точек.

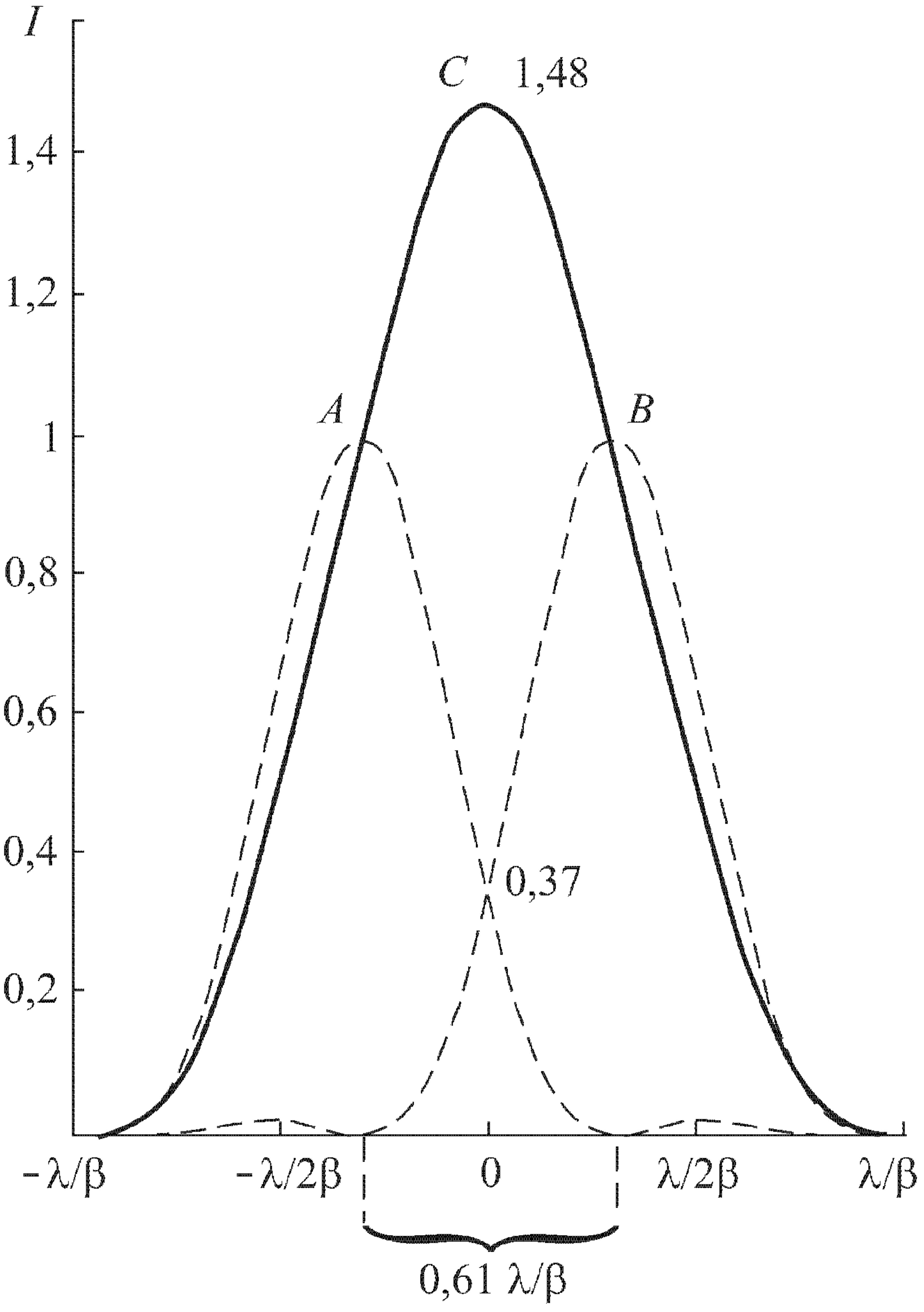


Рис. 214

Легко теперь сообразить, что следует ожидать при освещении объекта лучами всевозможных направлений от протяженного источника света. Нормально падающим лучам соответствует наименьшая разрешающая способность и наибольшее разрешаемое расстояние, превосходящее соответствующее расстояние для некогерентных светящихся точек. При возрастании угла наклона лучей разрешаемое расстояние уменьшается. При угле наклона, когда разность фаз Δ становится равной $\pi/2$, оно такое же, что и для некогерентных источников. При дальнейшем увеличении наклона разрешаемое расстояние продолжает монотонно уменьшаться и при скользящем падении становится почти вдвое

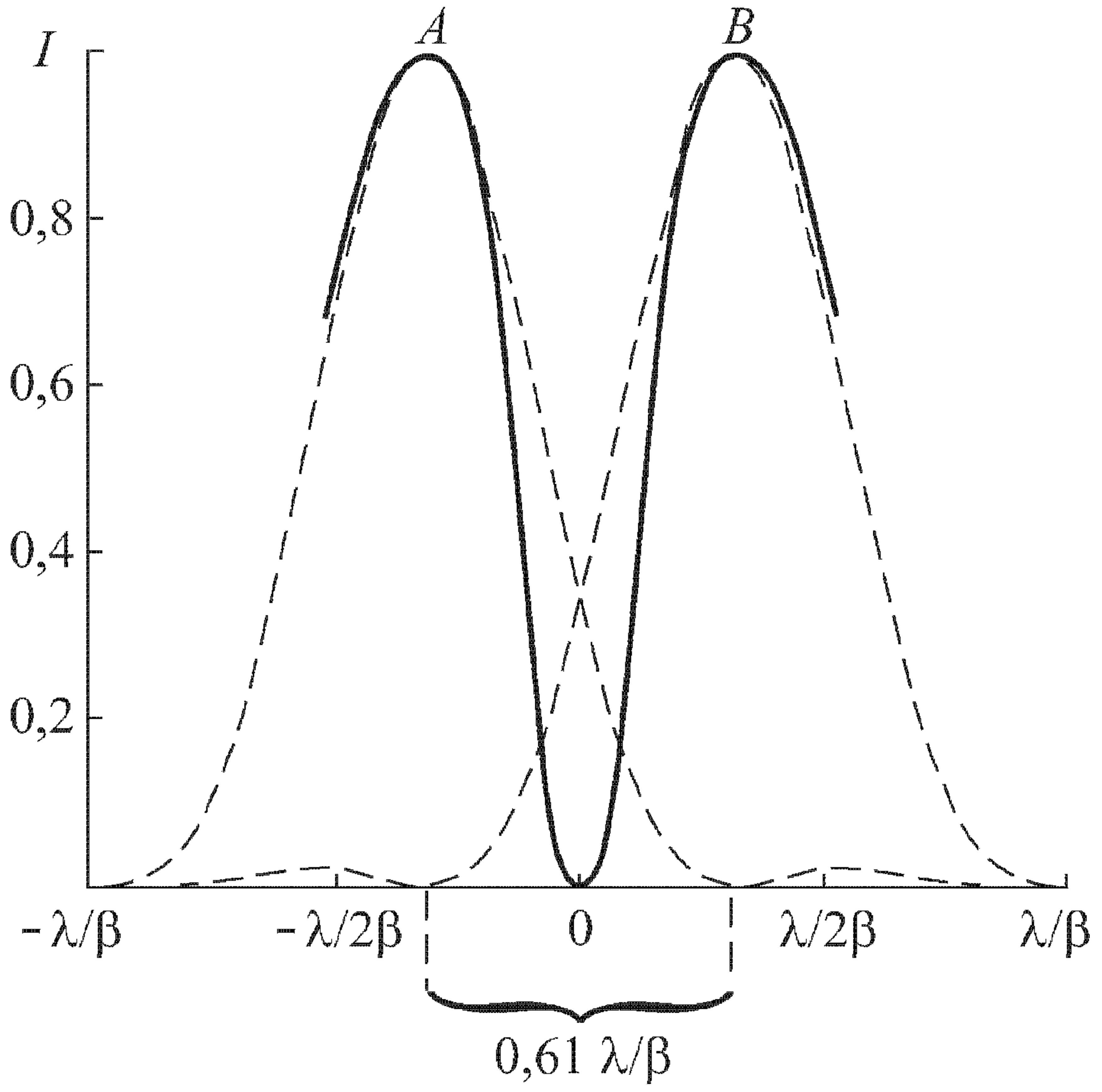


Рис. 215

меньше соответствующего расстояния для некогерентных источников. Естественно поэтому ожидать, и это подтверждается расчетом, что при освещении широкими пучками лучей получится такое же распределение интенсивности света в дифракционной картине и такой же критерий разрешения, что и при некогерентных источниках света.

Таким образом, разрешаемое расстояние меняется с изменением условий освещения и должно рассматриваться как ориентировочное. При этом надо еще раз подчеркнуть, что полученные результаты относятся к визуальным наблюдениям. Применение других физических методов исследования может понизить разрешаемый предел.

4. В случае телескопа рассматриваемые объекты, например компоненты двойной звезды, всегда излучают *некогерентно*. Для телескопа интерес представляет не линейное, а *угловое разрешаемое расстояние*. Если ϑ — угловое расстояние между рассматриваемыми звездами (рис. 216), то расстояние между центрами соответствующих кружков Эйри в фокальной плоскости будет $x = S_1'S_2' = f\vartheta$, где f — фокусное расстояние объектива.

Согласно критерию Рэлея, для разрешения звезд это расстояние должно быть не меньше $0{,}61\lambda/\beta$, т.е. $f\vartheta \geqslant 0{,}61\lambda/\beta$. Учитывая, что для небольших углов $f\beta = D/2$, где D — диаметр объектива, отсюда получим

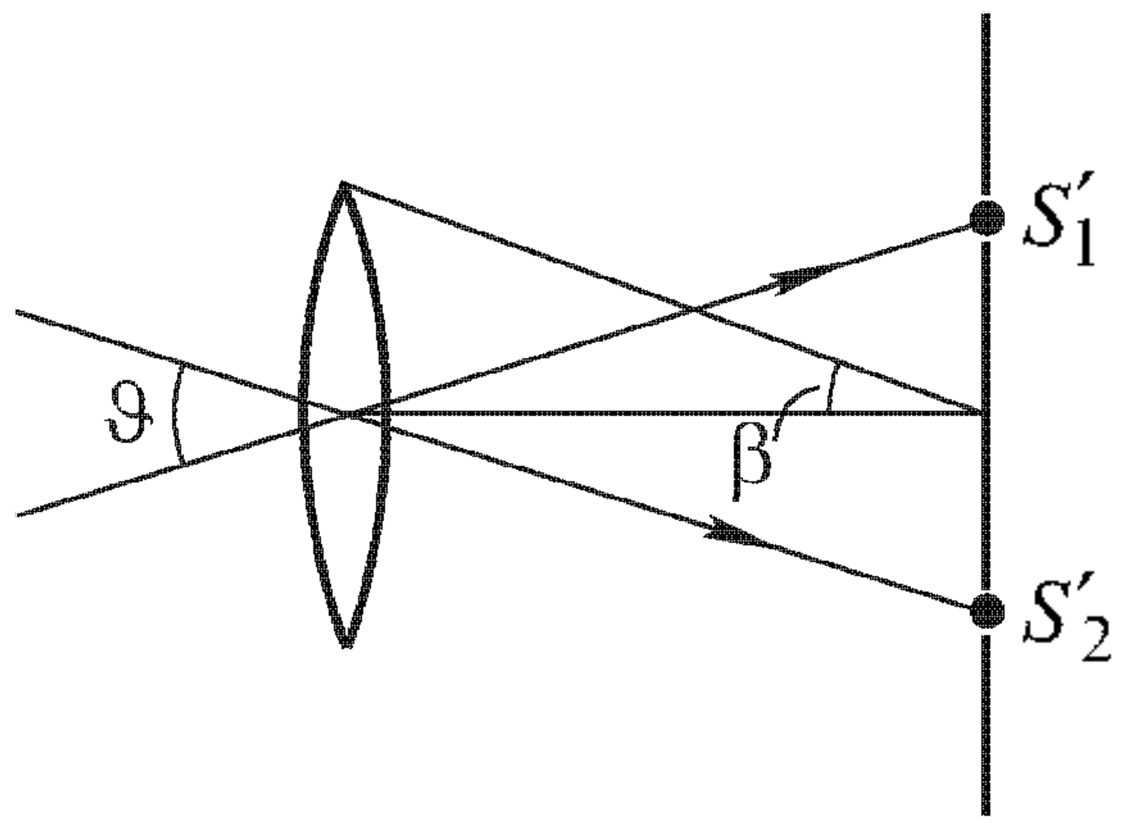


Рис. 216

$$\vartheta \geqslant 1{,}22\frac{\lambda}{D}. \qquad (56.1)$$

Разрешаемое угловое расстояние равно

$$\vartheta_{\text{мин}} = 1{,}22\frac{\lambda}{D}. \qquad (56.2)$$

Глаз при рассматривании удаленных предметов действует принципиально так же, как объектив телескопа. Поэтому формулы (56.1) и (56.2) применимы и к глазу. Роль D играет диаметр зрачка глаза d. Полагая $d = 4$ мм, $\lambda = 550$ нм, находим для разрешаемого расстояния человеческого глаза

$$\vartheta_{\text{гл}} = 1{,}22\frac{\lambda}{d} = 1{,}67 \cdot 10^{-4} \text{ рад} = 35''. \qquad (56.3)$$

Этот результат удивительно хорошо согласуется с физиологической оценкой разрешающей способности глаза (см. § 21, п. 7). Достойна восхищения способность живого организма приспосабливаться к окружающим условиям и в процессе эволюции достигать максимума того, что принципиально допускается законами природы.

Таким образом, разрешающая способность телескопа пропорциональна диаметру его объектива. Крупнейший в мире действующий телескоп-рефлектор (см. § 24) имеет диаметр параболического зеркала $D = 6$ м. Теоретическая разрешающая сила его превосходит разрешающую силу глаза в $6000/4 = 1500$ раз. Теоретическое разрешаемое расстояние составляет $35''/1500 = 0{,}023''$. Для второго по величине телескопа-рефлектора Маунт-Паломарской обсерватории с диаметром зеркала 5 м теоретическое разрешаемое расстояние равно $0{,}028''$. Разрешающей способности таких гигантских телескопов достаточно, чтобы изображения звезд с наибольшими угловыми размерами получались уже в виде дисков, подобно изображениям планет.

Однако наличие земной атмосферы не позволяет полностью использовать разрешающую способность телескопа. Нерегулярные процессы в атмосфере, сопровождающиеся изменениями показателя преломления на пути световых лучей, портят изображения и снижают реальную разрешающую способность телескопов до величин порядка $1''$ — $0{,}5''$, причем особенно сильно это снижение сказывается на больших телескопах. Такая разрешающая способность может быть достигнута с помощью много меньших объективов. Большие телескопы строятся не с целью повышения разрешающей способности, а для *увеличения количества света*, поступающего в телескоп от наблюдаемых небесных объектов. Оно пропорционально площади отверстия объектива. Поэтому с помощью больших телескопов можно обнаружить и сфотографировать более слабые небесные объекты, чем с помощью малых. Для повышения же разрешающей способности астрономических телескопов необходимо исключить вредное влияние атмосферы. Большие перспективы открываются перед телескопами, устанавливаемыми на космических кораблях, а еще лучше — на Луне.

5. Увеличение телескопа должно быть согласовано с разрешающей способностью его объектива. Допустим, что угловой размер предмета равен минимальному углу $\vartheta_{\text{мин}}$, разрешаемому объективом телескопа. Если угловое увеличение телескопа равно N, то в телескоп предмет будет виден под углом $N\vartheta_{\text{мин}}$. Для полного использования разрешающей способности телескопа необходимо, чтобы этот угол разрешался глазом, т.е. должно быть $N\vartheta_{\text{мин}} \geqslant \vartheta_{\text{гл}}$. Подставляя сюда значения $\vartheta_{\text{мин}}$ и $\vartheta_{\text{гл}}$ из формул (56.2) и (56.3), получим

$$N \geqslant \frac{D}{d} \equiv N_{\text{норм}}, \tag{56.4}$$

где $N_{\text{норм}}$ — нормальное увеличение телескопа (см. § 23, п. 4).

В § 11 (п. 10) было показано, что $N = D/h$, т.е. увеличение телескопа равно отношению ширины D падающего параллельного пучка лучей к ширине h выходящего пучка. Когда увеличение меньше нормального, т.е. $D/H < D/d$, то $h > d$. Значит, не весь пучок параллельных лучей, падающий на объектив, попадает в зрачок глаза, а только часть его. Действующей является только центральная часть объектива, диаметр которой меньше D. Телескоп действует так, как если бы его объектив и разрешающая способность уменьшились. При нормальном увеличении $h = d$, разрешающая способность объектива согласована с разрешающей способностью глаза. Наконец, когда увеличение больше нормального, т.е. $D/h > D/d$, то $h < d$. В этом случае глаз действует так, как если бы его зрачок сузился, а разрешающая способность понизилась.

Применение увеличений больше нормального не может выявить новые детали рассматриваемого объекта. Увеличения, значительно превосходящие нормальное, приносят только вред, поскольку чрезмерное сужение выходящего пучка может внести в изображение значительные дифракционные искажения. Однако по физиологическим причинам иногда бывает полезно брать увеличения, превосходящие нормальное в 2–4 раза. Рассматривая детали на пределе разрешающей способности, глаз работает с напряжением и быстро утомляется, его чувствительность и разрешающая способность понижаются. Применение увеличений, несколько превышающих нормальное, не выявляет новых подробностей в строении рассматриваемых объектов, но позволяет рассматривать их с меньшим напряжением. К таким же выводам мы пришли в § 23 из фотометрических соображений.

6. Вопрос о разрешающей способности микроскопа решается так же, как и для телескопа. Критерий разрешения Рэлея в этом случае требует

$$l' \geqslant 0{,}61\frac{\lambda}{\beta}, \tag{56.5}$$

где l' — расстояние между центрами кружков Эйри для рассматриваемых двух светящихся точек (предполагается, что апертурный угол со стороны изображения 2β мал). Чтобы перейти к линейным размерам самого объекта, следует воспользоваться условием синусов Аббе, которое должно выполняться для объектива всякого микроскопа. При малых углах β (рис. 217) оно может быть написано в виде $ln\sin\alpha = l'\beta$, где n — показатель преломления пространства объектов, а 2α — апертурный угол со стороны этого пространства. Исключая β и l', приходим к соотношению

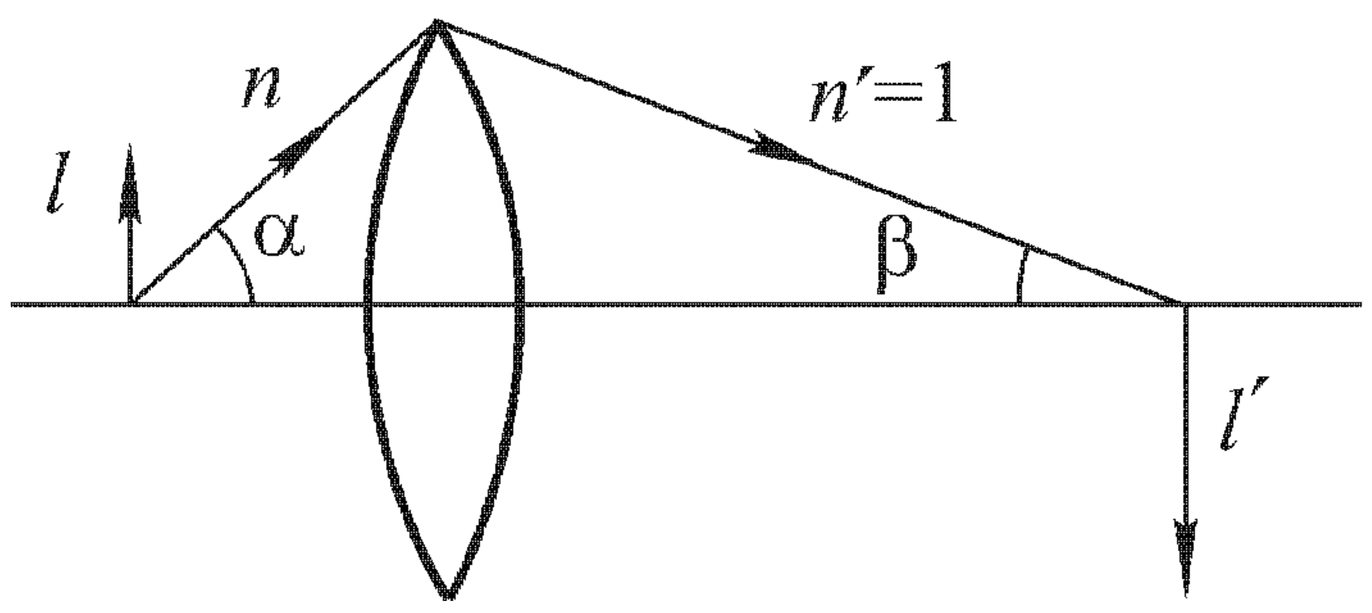


Рис. 217

$$l \geqslant 0{,}61\frac{\lambda}{n\sin\alpha}, \tag{56.6}$$

которое и определяет предел разрешения микроскопа. Минимальное разрешаемое расстояние (с точностью до небольшого численного коэффициента) равно

$$l_{\text{мин}} = 0{,}61\frac{\lambda}{n\sin\alpha}. \tag{56.7}$$

Этот предел разрешения определяется *волновой природой света* и не может быть превзойден никакими техническими усовершенствованиями микроскопа. При прочих равных условиях предел разрешения тем меньше, чем короче длина волны. Наличие в знаменателе показателя преломления n объясняется тем, что разрешаемое расстояние может непосредственно зависеть от длины волны не в вакууме, а в той среде, *где находится объект*, т.е. от величины λ/n.

Нетрудно проверить, что при малых апертурах формула (56.7) переходит в формулу (56.2), определяющую разрешающую способность телескопа. В этом случае обе формулы отличаются одна от другой только по форме. В формуле (56.7) разрешаемое расстояние выражено в линейных единицах, а в формуле (56.2) — в угловых.

Для повышения разрешающей способности микроскопа можно идти двумя путями: 1) уменьшать длину волны (переход к ультрафиолету); 2) увеличивать числовую апертуру объектива микроскопа $n\sin\alpha$. Таким образом, угол α должен быть как можно больше. В лучших современных объективах он практически достиг своего теоретического предела $\alpha = \pi/2$. Для повышения числовой апертуры применяют также иммерсию, т.е. жидкость с возможно высоким показателем преломления, заполняющую пространство между покровным стеклом и фронтальной линзой объектива (см. § 18, п. 4). Если принять $n \approx 1{,}5$, то максимальное значение числовой апертуры будет $n\sin\alpha \approx 1{,}5$. Применение иммерсии, таким образом, позволяет снизить разрешаемый предел примерно в полтора раза, т.е. довести его до величины $l \approx 0{,}61\lambda/1{,}5 \approx 0{,}4\lambda$. Детали объекта, размеры которых меньше примерно $0{,}4\lambda$, принципиально не могут быть выявлены с помощью микроскопа. Ни при каком увеличении нельзя определить форму объекта, рассматривая его в микроскоп, если размеры самого объекта меньше приблизительно $0{,}4\lambda$. Разумеется, с помощью микроскопа можно обнаружить объекты и много меньших размеров, если только они светятся достаточно ярко (ультрамикроскоп). Нельзя определить только их форму.

Радикальное повышение разрешающей способности достигается в электронном микроскопе, где роль световых лучей выполняют электроны, фокусируемые электрическими и магнитными линзами. Электроны ведут себя подобно волнам с длиной волны

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m\mathcal{E}}},$$

где h — постоянная Планка, m — масса, p — импульс, $\mathcal{E}$ — энергия электрона. Эти волны называются *волнами де Бройля* (р. 1892). При $\mathcal{E} = 10\,000$ эВ приведенная формула дает $\lambda = 0{,}0122$ нм. Такая длина волны на порядок меньше размеров атома. Однако такое разрешение в электронном микроскопе не достигается, так как для уменьшения геометрических аберраций в электронных микроскопах приходится пользоваться узкими параксиальными пучками с малыми апертурами. Но даже в таких пучках электронный микроскоп позволяет разрешать детали, превосходящие размеры атомов всего в несколько раз.

7. Рассмотрим в заключение вопрос о рациональном увеличении микроскопа. Напомним, что увеличением микроскопа (лупы) называется отношение угла, под которым виден объект в микроскоп, к тому углу, под которым он был бы виден невооруженным глазом, если бы был помещен на расстоянии ясного зрения. Пусть $l_{\text{мин}}$ — минимальное расстояние, разрешаемое микроскопом. Невооруженным глазом с расстояния ясного зрения оно видно под углом $\vartheta = l_{\text{мин}}/L$. В микроскоп то же расстояние видно под углом $\vartheta' = N\vartheta$. Для разрешения необходимо, чтобы угол ϑ' был не меньше минимального угла, разрешаемого глазом, т.е. $\vartheta' = N\vartheta \geqslant \vartheta_{\text{гл}}$. Отсюда, используя выражения (56.3) и (56.7), найдем

$$N \geqslant \frac{2Ln\sin\alpha}{d}. \tag{56.8}$$

Знаку равенства соответствует нормальное увеличение

$$N_{\text{норм}} = \frac{2Ln\sin\alpha}{d}. \tag{56.9}$$

Как и в случае телескопа, нормальное увеличение микроскопа есть наименьшее увеличение, при котором может быть использована вся разрешающая способность объектива. О целесообразности работать при увеличениях больше нормального для микроскопа можно повторить без всяких изменений все, что выше было сказано относительно телескопа.

Максимальная числовая апертура, как указывалось выше, для сухих систем имеет своим пределом единицу. Для иммерсионных систем этот предел равен примерно 1,5. Диаметр зрачка глаза d примем равным 2 мм. Тогда для нормального глаза ($L = 25$ см) получатся следующие предельные значения нормальных увеличений:

для сухих систем $N_{\text{норм}} = 250$,
для иммерсионных систем $N_{\text{норм}} = 375$.

По физиологическим соображениям имеет смысл переходить к несколько большим увеличениям. Однако бессмысленно строить микроскопы с увеличением больше, чем в 1000–1500 раз.

§ 57. Теория и демонстрационные опыты Аббе

1. Впервые предел разрешения объектива микроскопа был найден Гельмгольцем в 1874 г. Гельмгольц рассматривал *самосветящиеся объекты*. Примерно в то же время и независимо от Гельмгольца вопрос о разрешающей способности микроскопа был разобран Аббе, но для случая *освещаемых объектов*. Именно этот случай встречается в практике микроскопии. От освещения очень сильно зависит качество изображения в микроскопе. Этот вопрос подвергся глубокому изучению Д.С. Рождественским (1876–1940) в последние годы его жизни. Ниже излагается метод Аббе с некоторыми изменениями и затрагиваются дополнительные вопросы, связанные с этим методом.

Допустим сначала, что объектом является достаточно большая (бесконечная) плоская дифракционная решетка, освещаемая параллельным пучком монохроматических лучей. Свет, прошедший через решетку, будет состоять из дискретного ряда плоских волн (дифракционных пучков или спектров различных порядков), распространяющихся от решетки в различных направлениях. Пучки низких порядков будут однородными волнами, а пучки, порядок которых превосходит определенное значение, — неоднородными (см. § 52). Неоднородные волны, затухая на расстояниях порядка длины волны, в объектив не попадают и потому не влияют на изображение, даваемое им.

2. Допустим, что перед объективом поставлена диафрагма, пропускающая спектры определенных порядков. Если диафрагма пропускает пучок одного лишь нулевого порядка, то прошедший свет будет состоять только из одной однородной плоской волны, как если бы на месте дифракционной решетки стояла прозрачная идеально плоскопараллельная пластинка. Никакой информации о структуре объекта такой свет не несет. В плоскости изображения, сопряженной с плоскостью решетки, получится равномерно освещенное поле.

Возьмем теперь такую диафрагму, чтобы она пропускала только два спектра соседних порядков, например m-го и $(m+1)$-го (рис. 218). За решеткой эффективными останутся только две однородные волны, распространяющиеся в различных направлениях:

$$E_m = a_m \cos(\omega t - \mathbf{k}_m \mathbf{r}),$$
$$E_{m+1} = a_{m+1} \cos(\omega t - \mathbf{k}_{m+1} \mathbf{r}),$$

интерферирующие между собой. Плоскость решетки примем за координатную плоскость XY, направив ось X перпендикулярно к ее штрихам, а ось Z — в сторону распространения падающего света. Интерференционные полосы во всякой плоскости $z = \mathrm{const}$, параллельной плоскости решетки, будут параллельны ее штрихам. Найдем расстояние Δx между ними. Разность фаз между рассматриваемыми волнами в указанной плоскости представляется выражением

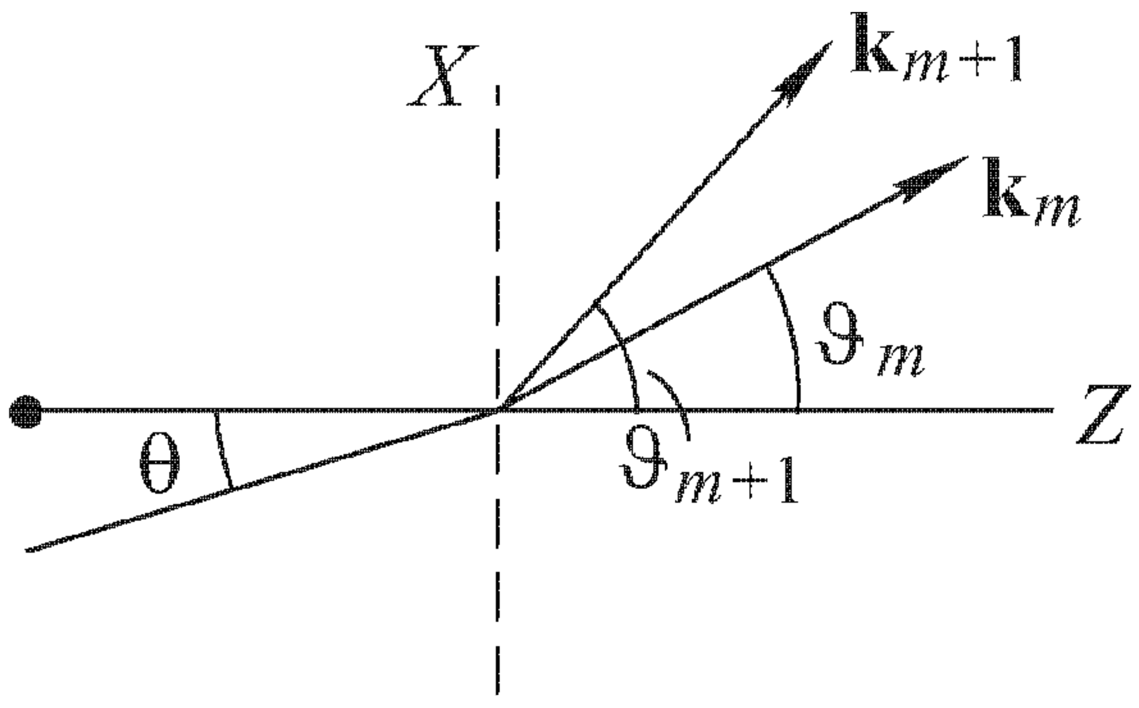


Рис. 218

$$\varphi = (\mathbf{k}_{m+1} - \mathbf{k}_m)\mathbf{r} = (k_{m+1,x} - k_{m,x})x + (k_{m+1,z} - k_{m,z})z.$$

При изменении x на Δx она изменится на $\Delta\varphi = (k_{m+1,x} - k_{m,x})\Delta x$. Интенсивность света будет периодически повторяться, когда $\Delta\varphi = 2\pi, 4\pi, \ldots$ Ширина интерференционной полосы Δx найдется, если положить $\Delta\varphi = 2\pi$. Направления на дифракционные максимумы рассматриваемых порядков определяются формулами

$$d(\sin\vartheta_m - \sin\theta) = m\lambda, \quad d(\sin\vartheta_{m+1} - \sin\theta) = (m+1)\lambda,$$

где θ — угол, под которым свет падает на решетку. Отсюда $d(\sin\vartheta_{m+1} - \sin\vartheta_m) = \lambda$. Но $k_{mx} = (2\pi/\lambda)\sin\vartheta_m$, $k_{m+1,x} = (2\pi/\lambda)\sin\vartheta_{m+1}$, так что

$$k_{m+1,x} - k_{mx} = (2\pi/\lambda)(\sin\vartheta_{m+1} - \sin\vartheta_m) = 2\pi/d.$$

Следовательно, ширина интерференционной полосы будет $\Delta x = 2\pi : (2\pi/d) = d$, независимо от направления освещающего пучка лучей. В частности, такая система интерференционных полос возникнет на выходе самой решетки, т.е. в плоскости $z = 0$. Объектив даст изображение этой системы полос в плоскости изображения, сопряженной с плоскостью решетки (см. § 27, п. 6). Такое изображение передает основной характер структуры решетки — ее периодичность с основным периодом d. Распределение интенсивности в плоскости изображения будет синусоидальным, т.е. таким же, какое получилось бы от синусоидальной решетки Рэлея. Чтобы изображение передавало более мелкие детали структуры объекта, необходимо расширить диафрагму. Тогда за решеткой появятся плоские дифрагированные волны, порядки которых отличаются друг от друга не только на единицу, но также на 2, 3, ... Рассуждая как выше, нетрудно показать, что интерференция двух волн, порядки которых отличаются на Δm, даст интерференционные полосы с шириной $\Delta x = d/\Delta m$.

Структуру решетки можно характеризовать разложением ее пропускаемости в ряд Фурье. Полностью подобное изображение со всеми деталями получилось бы, если бы интенсивность света в плоскости изображения представлялась рядом Фурье с теми же коэффициентами. Однако этого никогда не бывает. Из ряда Фурье выпадают компоненты, соответствующие неоднородным волнам и тем однородным волнам, которые задерживаются диафрагмой. Чем больше дифрагированных волн различных порядков проходит через диафрагму, тем совершеннее получается изображение.

3. Если при определении разрешающей способности довольствоваться наименее совершенным изображением, которое правильно передавало бы только периодическую структуру решетки с периодом d, то разрешающую способность объектива микроскопа можно определять следующим образом. Пусть решетка освещается нормально падающими лучами света (рис. 219 *а*). Тогда для разрешения необходимо, чтобы наряду с прямо прошедшим пучком света в объектив попали также дифракционные пучки первого и минус первого порядков. Если объектив иммерсионный с показателем иммерсии n, то минимальный период решетки, при котором выполняется это условие, будет $d\sin\alpha = \lambda/n$, где λ — длина волны в вакууме. Отсюда и находится минимальное расстояние, разрешаемое объективом микроскопа:

$$l_{\text{мин}} = \frac{\lambda}{n\sin\alpha}. \tag{57.1}$$

Полученный результат отличается от (56.7) только несущественным числовым коэффициентом. Различие в коэффициентах связано с некоторым различием критериев разрешения, которые использовались при выводе формул (56.7) и (57.1), а также с тем, что формула (56.7) относится к разрешению точечных, а (57.1) — линейных объектов.

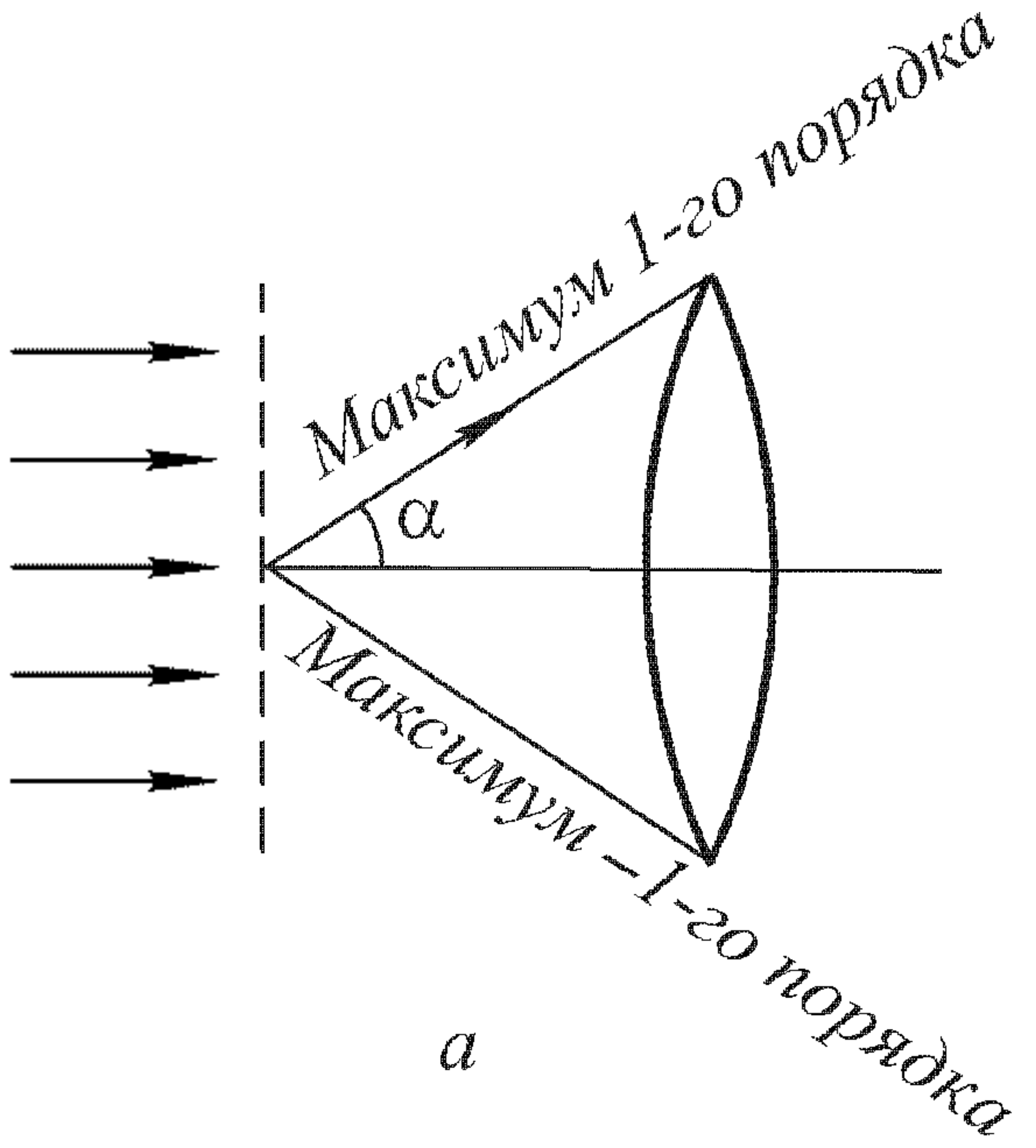


Рис. 219

Разрешаемый предел (57.1) можно понизить за счет применения наклонного освещения. При нормальном падении освещающего пучка за решеткой интерферируют три пучка. При интерференции пучков нулевого и одного из первых порядков в изображении выявляется структура с основным периодом d. Такая же картина интерференции получается при наложении нулевого пучка с другим пучком первого порядка. Второй пучок первого порядка просто усиливает действие другого пучка того же порядка. Кроме того, оба пучка первого порядка интерферируют между собой, в результате чего на интерференционную картину накладываются новые, более слабые, интерференционные полосы, расстояние между которыми вдвое меньше, т.е. равно $d/2$. В этом проявляется действие первых гармоник в разложении пропускаемости решетки в ряд Фурье. Не меняя основного периода в изображении решетки, они несколько усиливают его контраст. Но для выявления структуры с основным периодом d, как выяснено выше, достаточно *двух* соседних пучков, например, нулевого и одного из пучков первого порядка. Максимально возможное разрешение получится, если освещающий пучок направить под углом α к решетке, как указано на рис. 219 *б*. Тогда будет $2d\sin\alpha = \lambda/n$ и, следовательно,

$$l_{\text{мин}} = \frac{1}{2}\frac{\lambda}{n\sin\alpha}, \tag{57.2}$$

что вдвое меньше величины (57.1).

4. Таким образом, разрешающую способность микроскопа можно повысить изменением направления освещающих лучей. Однако влияние способа освещения на изображение в микроскопе этим не ограничивается. Оно более существенно. Если бы освещение производилось пучком параллельных лучей, то, как было выяснено выше, интерференционная картина за решеткой в каждой параллельной ей плоскости П имела бы вид параллельных полос с расстоянием между ними d. Такие полосы *не локализованы*. Изображение, видимое в микроскоп, имело бы один и тот же вид независимо от того, куда сфокусирован микроскоп: на плоскость решетки или на любую плоскость, ей параллельную. Для устранения этого вредного обстоятельства в микроскопе объекты освещаются пучками лучей различных направлений с широкими апертурами. Интерференционные полосы, даваемые параллельными пучками различных направлений, смещены относительно друг друга вдоль оси X. Поэтому во всех плоскостях П происходит гашение интерференционных полос. Исключение составляет только случай, когда плоскость П совпадает с плоскостью самой решетки. В этом случае все интерференционные полосы, создаваемые освещающими пучками различных направлений, пространственно совпадают между собой. Поэтому изображение объекта в микроскопе может возникнуть только тогда, когда он сфокусирован на плоскость самого объекта. В этом основная причина использования широких апертур при освещении объектов в микроскопе.

5. В качестве объекта была взята дифракционная решетка лишь с целью упрощения рассуждений. Это упрощение не принципиальное. Если взять объект произвольной формы и осветить его параллельным пучком лучей, то за объектом возникнут дифрагированные пучки различных направлений с максимумами и минимумами. Угол дифракционной расходимости ϑ, определяющий направление на первый дифракционный минимум, определяется формулой $nl\sin\vartheta \sim \lambda$, где l — линейный размер объекта. Если диафрагма пропускает лишь незначительную часть центрального дифракционного пучка, то изображение в микроскопе получится таким же, как и в случае точечного объекта, излучающего сферические волны равномерно по всем направлениям. Это был бы дифракционный кружок Эйри, совершенно не зависящий от формы объекта, и никакого разрешения не получилось бы. Минимальные размеры объекта $l_{\text{мин}}$, при которых возможно разрешение, определяются условием $\vartheta \sim \alpha$, т.е.

$$l_{\text{мин}} \sim \frac{\lambda}{n \sin \alpha}. \tag{57.3}$$

Мы снова пришли к формуле вида (57.1) или (57.2).

В приведенном рассуждении не столь существенно, куда попадает свет после рассеяния на объекте. Вместо объектива можно, например, взять фотопластинку и произвести на ней голографическую запись предмета. Из предыдущих рассуждений следует, что минимальный размер объекта, который может быть записан таким способом, определяется формулой

$$l_{\text{мин}} \sim \lambda/\alpha, \tag{57.4}$$

где α — угловой размер голограммы, если ее рассматривать из точки нахождения предмета во время голографирования (предполагается, что он невелик).

6. Сам Аббе представил изложенную теорию в следующей форме. Каждый пучок параллельных лучей, получившийся в результате дифракции на объекте AB (например, дифракционной решетке), собирается объективом в точке фокальной плоскости (рис. 220). Дифракционную картину в фокальной плоскости, состоящую из совокупности таких точек: C_0, C_1, C_{-1}, C_2, C_{-2}, ..., Аббе назвал *первичным изображением* объекта. Если известно первичное изображение, то с помощью принципа Гюйгенса–Френеля можно рассчитать световое поле во всем пространстве за фокальной плоскостью. В частности, изображение объекта есть не что иное, как интерференционная картина, получающаяся в плоскости изображения в результате интерференции вторичных волн Гюйгенса, исходящих из различных точек первичного изображения. Эту интерференционную картину Аббе назвал *вторичным изображением*.

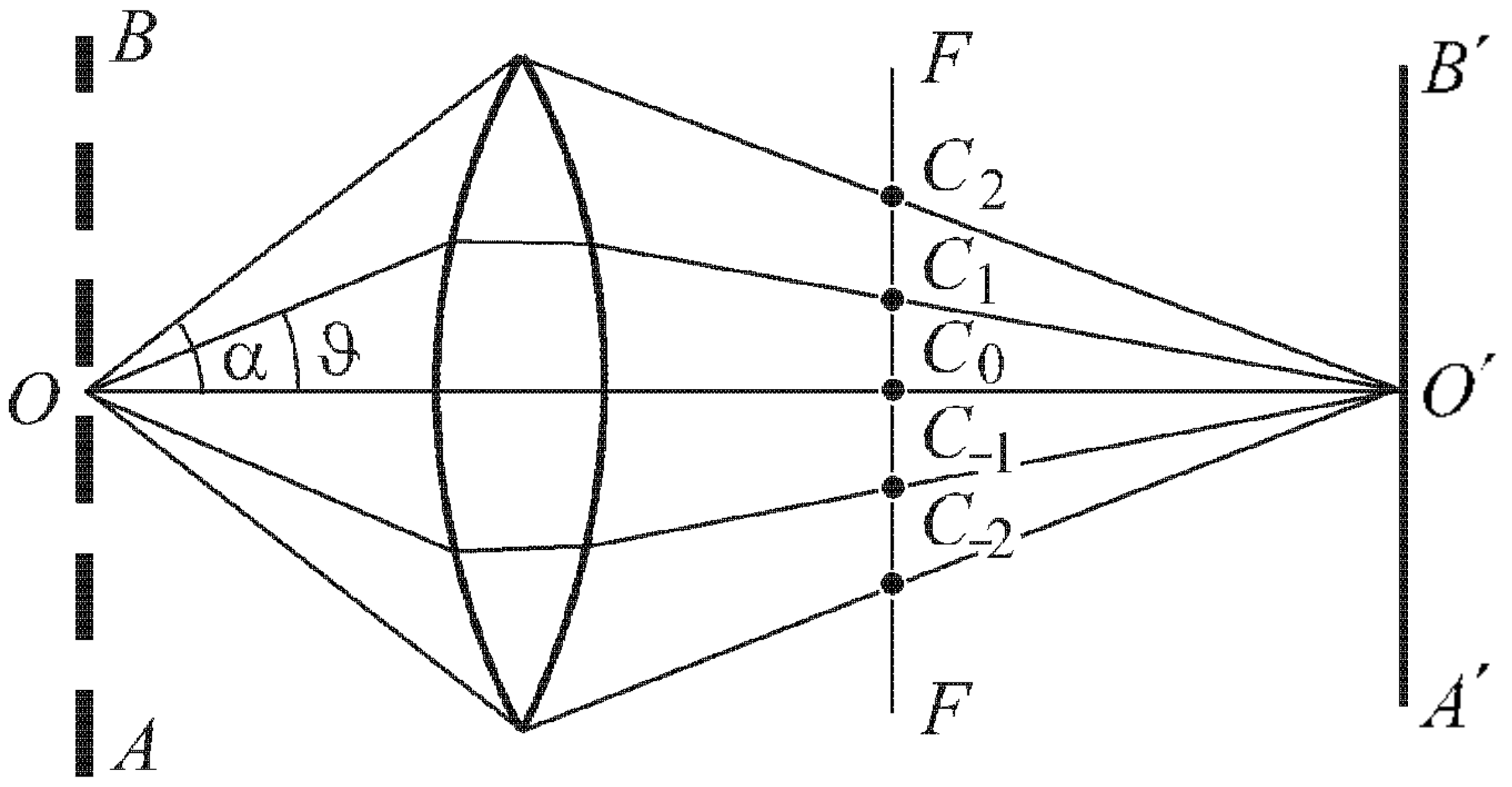


Рис. 220

Если диафрагма объектива пропускает один только центральный дифрагированный пучок, то первичное изображение будет состоять только из одного центрального точечного максимума. Получится лишь одна вторичная сферическая волна, исходящая из этого центрального максимума, которой не с чем будет интерферировать. Вторичное изображение получится бесструктурным и будет представлять собой более или менее равномерно освещенное поле. Для появления какой-то структуры во вторичном изображении необходимо, чтобы первичное изображение состояло по крайней мере из двух точечных дифракционных максимумов. Но это условие лишь по форме отличается от того условия, из которого были выведены формулы (57.1) и (57.2).

Если закрыть некоторые дифракционные максимумы в фокальной плоскости, то вторичное изображение будет искажено. Характер искажения часто можно предсказать. Этим воспользовался Аббе для экспериментального подтверждения своей теории. Объектом служила грубая дифракционная решетка. Аббе помещал в фокальной плоскости другую решетку из проволок, которые закрывали дифракционные максимумы через один. Тогда расстояния между действующими (открытыми) максимумами возрастали вдвое. В изображении решетки наблюдалось удвоение штрихов. Все происходило так, как если бы экранирования не было, а объект был заменен более мелкой решеткой, расстояния между штрихами которой вдвое меньше.

Еще более интересные искажения наблюдаются в случае двумерной решетки, например, квадратной проволочной сетки. В этом случае дифракционная картина в фокальной плоскости состоит из светлых пятен, расположенных в узлах также квадратной сетки. Поместим в фокальной плоскости узкую щель, с помощью которой можно открывать прямолинейные ряды из таких пятен, а остальные пятна закрывать. Если щель горизонтальна и достаточно узка, то она выделит прямолинейный ряд максимумов, расположенных вдоль горизонтальной прямой (рис. 221 *а*). Такой ряд максимумов аналогичен дифракционной картине от одномерной решетки с вертикальными щелями. Поэтому оптическое изображение квадратной сетки при вве-

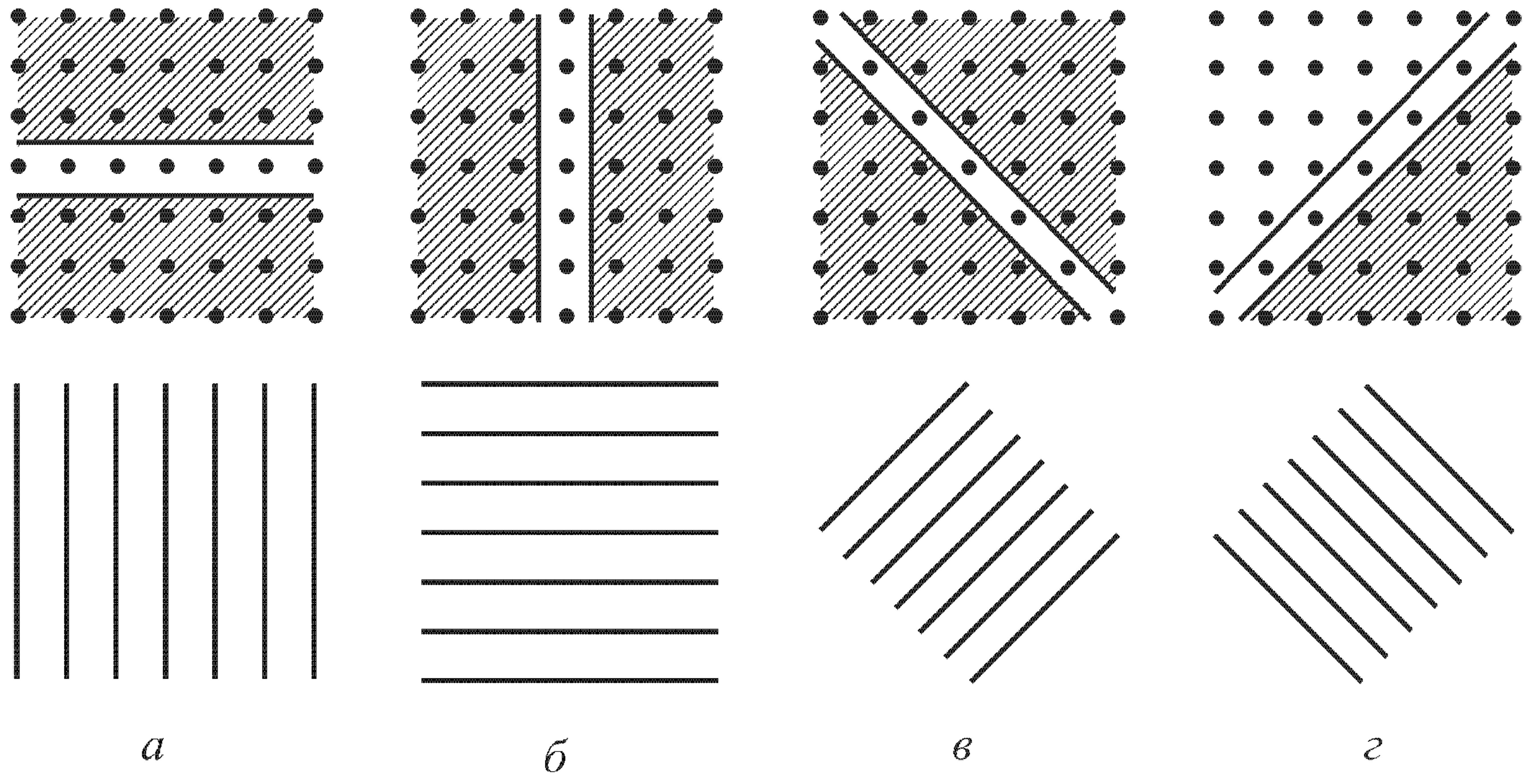


Рис. 221

дении горизонтальной щели перейдет в систему вертикальных полос. Если щель повернуть на 90° в вертикальное положение (рис. 221 *б*), то полосы сделаются горизонтальными. Если щель повернуть параллельно диагонали сетки (рис. 221 *в* и *г*), то она выделит прямолинейный ряд максимумов, параллельный той же диагонали, причем расстояния между максимумами увеличатся в $\sqrt{2}$ раз. В результате оптическое изображение сетки перейдет в систему наклонных полос, перпендикулярных к щели, а сами полосы сделаются в $\sqrt{2}$ раз у́же.

Д.С. Рождественский указал, что непосредственной причиной появления ложных структур в опытах Аббе является *дифракция света на экранирующей сетке.* В отсутствие последней объектив дает геометрически подобное изображение объекта. Например, сетка из параллельных проволок изображается в виде системы параллельных полос. При введении экранирующей сетки эти полосы остаются на прежних местах. Но на них накладывается дифракционная картина, возникающая при дифракции света на экранирующей сетке. Если экранирующая сетка состоит из проволок, параллельных проволокам сетки-объекта, то возникнут дифракционные полосы, параллельные полосам в прежнем изображении сетки-объекта. При надлежащем периоде экранирующей сетки дифракционные полосы расположатся посередине между прежними полосами. Произойдет удвоение полос в изображении, как если бы проволоки в сетке-объекте были натянуты вдвое чаще. Так же, хотя и не столь просто, объясняется и появление диагоналей, когда объектом является квадратная проволочная сетка.

Дифракция на экранирующей сетке происходит независимо от того, освещается ли объект посторонним светом или является самосветящимся. Поэтому и в случае самосвечения должны появляться такие же ложные структуры, как и при освещении. Это было теоретически предсказано Л.И. Мандельштамом (1879–1944) еще в 1911 г. Опыты с накаленными сетками, поставленные самим Мандельштамом, подтвердили это предсказание.

§ 58. Телескоп без объектива. Получение изображений с помощью малых отверстий

1. В принципе можно построить телескоп без объектива сколь угодно высокой разрешающей способности. Роль объектива может играть круглое отверстие. Следующее рассуждение, принадлежащее Рэлею, разъясняет идею такого телескопа. Допустим сначала, что в отверстие вставлен объектив. Лучи, идущие от какой-либо точки объекта к ее изображению вдоль оптической оси и по периферии, имеют различные геометрические длины. При отсутствии объектива различны и их оптические длины. Объектив должен скомпенсировать своей толщиной различие оптических длин всех лучей, чтобы они приходили в точку-изображение в одинаковых фазах. Но небольшое расхождение в фазах лучей мало сказывается на результате их интерференции. Например, если интерферируют два одинаковых луча в одинаковых фазах, то результирующая интенсивность превосходит интенсивность отдельного луча в 4 раза. Если же разность фаз составляет $\pi/4$, то она будет превосходить в $2+\sqrt{2} = 3{,}42$ раза, т.е. изменится незначительно. Поэтому разности фаз в $\pi/4$ или меньше можно не принимать во внимание. Допустим теперь, что фокусное расстояние объектива увеличивается все больше и больше. Различие в геометрических длинах центрального и крайнего периферийного лучей будет становиться все меньше и меньше. Когда это различие достигнет примерно $\lambda/8$, надобность в объективе отпадет. Действительно, в этом случае максимальная разность фаз между лучами, приходящими в точку-изображение от различных точек плоскости отверстия, не будет превышать $\pi/4$, и компенсация различия оптических длин становится практически не нужной. Отверстие будет действовать как объектив телескопа. При увеличении фокусного расстояния f разрешающая способность объектива не меняется и все время определяется формулой (56.2). Она останется неизменной и в предельном случае, когда $f=\infty$, т.е. когда объектив совсем удален.

Разность геометрических длин крайнего периферийного и центрального лучей составляет $D^2/(8l)$, где D — диаметр отверстия, а l — длина телескопа, которую можно считать равной фокусному расстоянию. Для оценки l приравняем эту разность $\lambda/8$ и получим

$$l = D^2/\lambda. \tag{58.1}$$

Эта формула показывает, что о практической реализации телескопа без объектива не может быть и речи ввиду колоссальных размеров, которые должен был бы иметь такой телескоп. Например, при $D = 1$ м, $\lambda = 500$ нм $= 5\cdot10^{-7}$ м формула (58.1) дает $l = 2\cdot10^6$ м $= 2000$ км. При гигантских размерах телескоп обладал бы ничтожной светосилой: $(D/f)^2 = l\lambda/f^2 = \lambda/l \approx 10^{-13}$.

2. К идее безлинзового телескопа примыкает способ получения изображений и фотографирования при помощи камеры с малым отверстием (камеры-обскуры). Она отличается от безлинзового телескопа только масштабами и возможностью практического осуществления. Поэтому приводимые ниже рассуждения относятся не только к камере с малым отверстием, но дополняют и уточняют идею безлинзового телескопа.

При уменьшении размеров отверстия четкость изображения в камере сначала улучшается, а затем начинает ухудшаться из-за дифракции. Дифракция несущественна при больших отверстиях, а при малых отверстиях становится основным фактором, определяющим четкость изображения. Оптимальные размеры отверстия, при которых достигается наибольшая четкость, легко оценить с помощью следующих соображений. Пусть отверстие имеет форму круга радиуса R. Расстояние до фотографируемого объекта может считаться бесконечно большим по сравнению с глубиной камеры l. Если бы была применима геометрическая оптика, то светящаяся точка изобразилась бы кружком того же радиуса R. Из-за дифракции точка изобразится дифракционным кружком, радиус которого порядка $\lambda l/R$. Уменьшать размеры отверстия имеет смысл лишь до тех пор, пока дифракционные ошибки не превзойдут геометрические. Наилучшая четкость изображения достигается при таких размерах отверстия, когда эти ошибки примерно одинаковы, т.е. при выполнении условия $\lambda l/R \approx R$, или $R \approx \sqrt{l\lambda}$. Это значит, что размер отверстия должен быть порядка центральной френелевой зоны. Рэлей, более подробно исследовавший вопрос как теоретически, так и экспериментально, нашел для наивыгоднейшего радиуса отверстия:

$$R = 0{,}9\sqrt{\frac{ab\lambda}{a+b}}, \qquad (58.2)$$

где a и b — расстояния предмета и его изображения от отверстия.

ЗАДАЧИ

1. При наблюдении в телескоп с нормальным увеличением освещенность изображения звезды на сетчатке глаза в $\alpha = 10$ раз меньше освещенности дневного неба, рассматриваемого в тот же телескоп. Во сколько раз надо увеличить диаметр объектива для того, чтобы освещенность изображения звезды на сетчатке стала в $\beta = 10$ раз больше освещенности изображения неба, если вместе с объективом телескопа заменен и окуляр таким образом, что увеличение телескопа осталось нормальным?

Ответ. В $\sqrt{\alpha\beta} = 10$ раз.

2. В фокальной плоскости объектива телескопа помещена фотопластинка. Освещенность изображения звезды на фотопластинке в $\alpha = 10$ раз меньше освещенности дневного неба. Во сколько раз надо увеличить диаметр объектива, чтобы освещенность изображения звезды на фотопластинке стала в $\beta = 10$ раз больше освещенности изображения неба?

Ответ. В $\sqrt{\alpha\beta} = 10$ раз.

3. Ракету, удаляющуюся от Земли, фотографируют, помещая фотопленку в фокальной плоскости объектива астрономического телескопа. При диаметре объектива $D_1 = 80$ мм дифракционное изображение ракеты становится неразличимым на фоне изображения неба, когда ракета находится на расстоянии $l_1 = 2 \cdot 10^4$ км от Земли. На каком расстоянии от Земли l_2 удастся заметить ракету на фотографии, полученной с помощью объектива с диаметром $D_2 = 200$ мм при той же контрастной чувствительности пленки?

Ответ. $l_2 = l_1 D_2/D_1 = 5 \cdot 10^4$ км.

4. Как изменится разрешающая способность объектива телескопа, если его центральную часть закрыть круглым экраном, диаметр которого мало отличается от диаметра самого объектива?

Указание. При решении задачи круглые объектив и экран можно заменить квадратными.

Ответ. Разрешающая способность увеличится приблизительно в два раза.

5. Производится фотографирование удаленных предметов с помощью объектива телескопа на фотопластинке, помещенной в его фокальной плоскости. Полученный снимок с помощью окуляра того же телескопа проецируется на удаленный экран. Каково должно быть угловое увеличение телескопа, чтобы при этом была использована полностью разрешающая способность объектива телескопа? Изображение на экране рассматривается с того места, где установлен проекционный аппарат.

Ответ. $N \geqslant D/d$, где D — диаметр объектива, а d — зрачка глаза.

6. С какого расстояния r можно увидеть невооруженным глазом свет лазера, генерирующего в непрерывном режиме $P = 10$ кВт на частоте $\omega = 4 \cdot 10^{15}$ с$^{-1}$, если для формирования луча используется параболическое зеркало с диаметром $D = 5$ м? Глаз видит источник, если в зрачок диаметром $d = 5$ мм попадает в 1 с $n = 60$ квантов излучения, лежащего в зеленой части спектра.

Ответ. $r \approx \frac{Dd}{2}\sqrt{\frac{P}{hc\lambda n}} \approx 0{,}1$ светового года (h — постоянная Планка.)

7. Излучение лазера непрерывного действия на волне $\lambda = 500$ нм мощностью $P = 1$ Вт направляется на спутник с помощью телескопа, диаметр объектива которого равен $D = 30$ см. Свет, рассеянный спутником, улавливается другим таким же телескопом и фокусируется на фотоумножитель с пороговой чувствительностью $P_{\text{пор}} = 10^{-14}$ Вт. При каких расстояниях l до спутника отраженный сигнал может быть обнаружен, если поверхность спутника равномерно рассеивает падающий на него свет (по закону Ламберта)? Диаметр спутника $d = 20$ см.

Ответ. $l \lesssim \frac{D}{2}\sqrt[4]{\frac{Pd^2}{2P_{\text{пор}}\lambda^2}} \approx 250$ км.

8. В § 40 был описан способ Поля фотографирования предметов с помощью непрозрачного гладкого шара. Оценить минимальное угловое разрешаемое расстояние $\delta\varphi$ при фотографировании удаленных предметов этим способом.

Указание. Размер светлого кружка в центре геометрической тени от точечного источника можно оценить из условия, что первый дифракционный минимум получается, когда разность хода лучей от противоположных точек шара порядка длины волны.

Ответ. $\delta\varphi \approx \lambda/D$.

9. Существующие радиотелескопы и радиоустановки, предназначенные для изучения радиоизлучения Солнца и Галактики, обладают малой разрешающей способностью из-за больших длин волн радиоизлучения.

1) Найти минимальное разрешаемое угловое расстояние $\delta\varphi$ радиотелескопа с диаметром зеркала $d = 50$ м для длин волн $\lambda = 1$ м и $\lambda = 10$ см.

2) Для увеличения разрешающей способности предлагалось использовать дифракцию радиоволн от края Луны (см. задачу 2 к § 42). Оценить разрешающую способность этого метода для тех же волн в предположении, что край Луны действует как тонкий непрозрачный экран, ограниченный прямолинейным краем.

3) Оценить, каковы должны быть высоты h неровностей лунной поверхности, чтобы можно было пользоваться этим методом. Расстояние до Луны $b = 380\,000$ км.

Ответ. 1) $\delta\varphi \approx \lambda/d$. При $\lambda = 1$ м $\delta\varphi \approx 1^\circ$; при $\lambda = 10$ см $\delta\varphi \approx 7''$.

2) $\delta\varphi \approx \sqrt{\lambda/b}$. При $\lambda = 1$ м $\delta\varphi \approx 2'$; при $\lambda = 10$ см $\delta\varphi \approx 40''$.

3) $h \ll \sqrt{b\lambda}$. При $\lambda = 1$ м $\sqrt{b\lambda} = 19{,}5$ км; при $\lambda = 10$ см $\sqrt{b\lambda} = 6{,}2$ км.

10. Один из принципиально возможных (но практически не осуществимых) способов повышения разрешающей способности радиоустановок для изучения радиоизлучения космических масс состоит в том, чтобы использовать дифракционный максимум интенсивности радиоизлучения, получающийся в центре геометрической тени Луны от точечного источника. Оценить разрешающую способность этого метода, найти минимальное угловое разрешаемое расстояние для тех же длин волн, что и в предыдущей задаче.

Обсудить возможности использования рассматриваемого метода. Диаметр Луны $D = 3470$ км.

Ответ. $\delta\varphi \approx \lambda/D$. При $\lambda = 1$ м $\delta\varphi \approx 0{,}06''$. При $\lambda = 10$ см $\delta\varphi \approx 0{,}006''$.

Метод требует, чтобы источник радиоизлучения находился на прямой, соединяющей точку наблюдения с центром Луны. Кроме того, он предъявляет весьма жесткие требования к гладкости лунной поверхности и к отступлениям формы Луны от сферической. Высота неровностей поверхности по всей границе лунного диска должна быть мала по сравнению с $h = b\lambda/D$. Разность наибольшего и наименьшего диаметров лунного диска также не должна превышать этой величины. При $\lambda = 1$ м $h \approx 100$ м; при $\lambda = 10$ см $h \approx 10$ м. Эти жесткие требования исключают возможность использования рассматриваемого метода по крайней мере для радиоволн с длиной волны $\lambda < 100$ м.

11. С помощью объектива микроскопа получена микрофотография малого объекта (например, растительных клеток или бактерий) с линейным увеличением N. Тот же объектив был использован для проецирования полученной микрофотографии на удаленный экран. Каково должно быть минимальное значение увеличения N, чтобы полностью была использована разрешающая способность микроскопа? Диаметр апертурной диафрагмы объектива равен D, диаметр зрачка глаза d. Изображение на экране рассматривается с места нахождения объектива.

Ответ. $N \geqslant D/d$.

12. Каково должно быть фокусное расстояние f_2 окуляра микроскопа, чтобы была полностью использована разрешающая способность объектива? Числовая апертура объектива равна $n\sin\alpha$, фокусное расстояние объектива f_1, длина тубуса (трубы микроскопа) l. Длину тубуса можно считать равной расстоянию между объективом и плоскостью первого изображения (т.е. изображения, даваемого объективом).

Ответ. $f_2 \leqslant \dfrac{ld}{2f_1 n\sin\alpha}$, где d — диаметр зрачка глаза.

13. С помощью объектива телескопа с диаметром D и фокусным расстоянием f производится фотографирование удаленных объектов на мелкозернистой пластинке, помещенной в фокальной плоскости объектива. Полученное изображение рассматривается в микроскоп с числовой апертурой $n\sin\alpha$ и увеличением N. Каким условиям должны удовлетворять числовая апертура и увеличение микроскопа, чтобы полностью использовать разрешающую способность объектива телескопа?

Ответ. $n\sin\alpha \geqslant D/(2f)$, $N \geqslant DL/(fd)$, где L — расстояние ясного зрения, d — диаметр зрачка глаза.

14. Современные фотопластинки способны разрешать до $z = 10^4$ линий на сантиметр. Какую светосилу (т.е. отношение квадратов диаметра D и фокусного расстояния f) должен иметь объектив фотоаппарата, чтобы полностью использовать разрешающую способность пленки?

Ответ. $(D/f)^2 \geqslant z^2\lambda^2 \approx 0{,}25$.

§ 59. Фазовый контраст

1. В зависимости от вида изучаемых объектов при работе микроскопа следует различать два предельных случая. Одни объекты, называемые *абсорбционными*, в различных местах обладают различной прозрачностью. Такие объекты в основном влияют на *амплитуду* проходящего света. Другие объекты, называемые *рефракционными*, практически не поглощают света. Имея в различных местах различные толщины и показатели преломления, они влияют не на интенсивность, а на *фазу* проходящего света. Типичными примерами абсорбционного и рефракционного объектов могут служить амплитудная и фазовая дифракционные решетки.

Абсорбционные объекты дают *контрастные изображения* с хорошо выраженными границами между темными и светлыми частями. На них можно обнаружить все детали, которые способен разрешить микроскоп при заданной разрешающей способности. Напротив, изображения рефракционных объектов почти *лишены контраста*. В таких изображениях трудно, а часто и практически невозможно разрешить детали изучаемого объекта, хотя бы разрешающей способности микроскопа и было достаточно для этой цели. Причина такого различия между абсорбционными и рефракционными структурами состоит в том, что объектив микроскопа воспроизводит в плоскости изображения, а следовательно и на сетчатке глаза, то же распределение интенсивности светового поля, которое существует в плоскости объекта, а светочувствительные нервные окончания сетчатки реагируют именно на *интенсивность* световой волны, а не на ее фазу.

С рефракционными объектами постоянно приходится иметь дело в биологии при изучении хотя бы микроорганизмов. Биологические объекты в подавляющем большинстве случаев практически совершенно прозрачны в видимой области спектра. Отсутствие контраста в изображении затрудняет изучение таких объектов. Поэтому проблема контрастности изображения стоит в биологии особенно остро. Один из методов ее решения состоит в превращении рефракционных объектов в абсорбционные путем дифференциального окрашивания объекта. Однако такой метод не всегда возможен. Кроме того, он убивает живые организмы или по крайней мере нарушает их нормальную жизнедеятельность. Единственный метод изучения биологических объектов в естественных условиях состоит в том, чтобы воздействовать не на самый объект, а на *его изображение*. Это достигается в *методе фазового контраста*, предложенном Цернике (1888–1966) в 1934 г.

2. Идею метода фазового контраста проще всего выяснить на примере периодической структуры — одномерной дифракционной решетки. Различие между амплитудной и фазовой решетками с интересующей нас точки зрения по существу уже было выяснено в § 53. Остановимся на этом вопросе более подробно.

Отвлекаясь от поляризации, будем рассматривать свет как скалярное волновое поле и представлять световые колебания векторами на векторной диаграмме. Пусть свет падает нормально на поверхность решетки. Допустим сначала, что решетка амплитудная и состоит из чередующихся участков различной прозрачности, причем на участках I прозрачность больше, а на участках II — меньше. Для простоты (это несущественно для выяснения существа вопроса) предположим, что участки I и II имеют одинаковую ширину. Колебание на выходе участка I изобразится более длинной стрелкой **a**, чем колебание на выходе участка II, представляемое стрелкой **b** (рис. 222). Так как амплитудная решетка не вносит разности фаз между волнами, прошедшими через различные участки ее, то обе стрелки **a** и **b** будут направлены одинаково.

Допустим теперь, что решетка фазовая и геометрически подобна амплитудной, т.е. состоит из чередующихся участков той же ширины, влияющих на фазу, но не на амплитуду волны. Колебания на выходе всех участков изобразятся стрелками **A** и **B** одинаковой длины, но различно направленными (рис. 222, положение *a*). Поскольку существенна лишь относительная разность фаз между обоими колебаниями, стрелки $\mathbf{A}$ и $\mathbf{B}$ можно повернуть на один и тот же угол, ничего не меняя в физических условиях задачи. Поэтому, не нарушая общности, можно предположить, что биссектриса угла между векторами $\mathbf{A}$ и $\mathbf{B}$ горизонтальна, т.е. параллельна поверхности решетки. Разложим каждый из векторов $\mathbf{A}$ и $\mathbf{B}$ на горизонтальную и вертикальную составляющие: $\mathbf{A} = \mathbf{D} + \mathbf{c}$, $\mathbf{B} = \mathbf{D} - \mathbf{c}$ (рис. 222, положение *б*). Допустим теперь, что оба вектора $\mathbf{c}$ и $-\mathbf{c}$ повернуты на 90° в одном и том

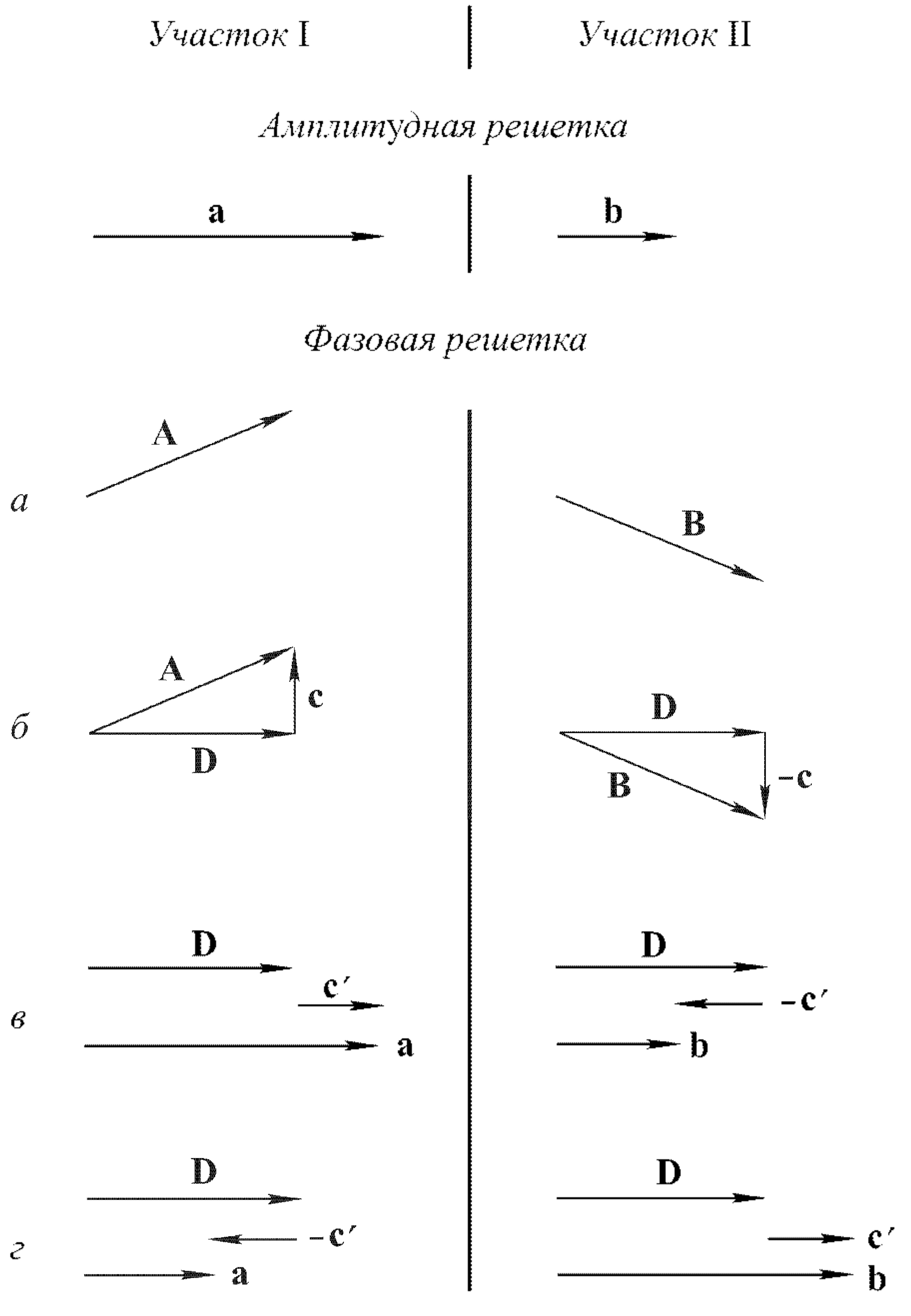


Рис. 222

же направлении в положении $\mathbf{c}'$ и $-\mathbf{c}'$ (рис. 222, положение *в*). Тогда на выходе участков I и II колебания представятся векторами $\mathbf{a} = \mathbf{D} + \mathbf{c}'$ и $\mathbf{b} = \mathbf{D} - \mathbf{c}'$, параллельными поверхности решетки. Поле на выходе фазовой решетки будет иметь такой же вид, что и на выходе амплитудной решетки. Поворот на 90° означает изменение фаз соответствующих колебаний на такой же угол. Таким образом, изменением фазы колебания на 90° можно превратить фазовую решетку в амплитудную. В этом и состоит идея метода фазового контраста.

При рассмотренном повороте по часовой стрелке вектор **a** получается длиннее вектора **b**. Это значит, что светлым местам в изображении амплитудной решетки будут соответствовать светлые же места в изображении фазовой решетки, а темным — темные (*позитивный фазовый контраст*). Если же векторы **c** и $-$**c** повернуть на 90° в противоположном направлении (рис. 222, положение *г*), то соотношение между длинами векторов **a** и **b**, а с ним и соответствие между светлыми и темными частями заменятся на противоположные (*негативный фазовый контраст*).

3. Чтобы повернуть векторы **c** и $-$**c**, сохраняя неизменным направление вектора **D**, надо прежде всего пространственно разделить волновые поля, представляемые этими векторами. Полное колебание на выходе решетки можно разложить на два колебания. Одно колебание имеет постоянную амплитуду на протяжении всей решетки и изображается постоянным вектором **D**. Оно дает в фокальной плоскости объектива центральный максимум нулевого порядка и не влияет на все остальные максимумы. Другое колебание представляется периодической функцией, которая равна $+$**c** на одних участках решетки и $-$**c** на соседних участках. Так как среднее по периоду решетки значение такой функции равно нулю, то такое колебание будет возбуждать только боковые максимумы, не оказывая никакого влияния на центральный максимум нулевого порядка. Таким образом, в фокальной плоскости объектива оба колебания окажутся пространственно разделенными. Одно концентрируется в центральном максимуме, другое распределяется по всем остальным — боковым — максимумам. Поставив на пути либо центрального максимума, либо всех боковых максимумов прозрачную плоскопараллельную пластинку надлежащей толщины, можно внести необходимую разность фаз в 90° и тем самым осуществить фазовый контраст. Такая пластинка называется *фазовой*.

4. До внесения фазовой пластинки энергия на участке I в условных единицах представляется выражением $A^2 = D^2 + c^2$. Такова же энергия на участке II. Полная энергия на обоих участках равна $2(D^2 + c^2)$. После поворота векторов **c** и $-$**c** на 90° энергии на участках I и II будут равны соответственно $(D + c)^2$ и $(D - c)^2$, а их сумма $2(D^2 + c^2)$. Таким образом, энергия не изменяется, а лишь перераспределяется между участками I и II. Этим перераспределением и объясняется просветление участков I и потемнение участков II.

§ 60. Измерение угловых диаметров звезд

1. Пусть перед объективом телескопа помещен экран с двумя круглыми отверстиями, расстояние между центрами которых равно D (рис. 223). Направив телескоп на звезду, закроем сначала правое отверстие. В результате дифракции на открытом левом отверстии в фокальной плоскости объектива получится система дифракционных колец. Их положение и размеры будут зависеть только от диаметра отверстия, но не от его положения в плоскости экрана. Поэтому, если закрыть левое отверстие и открыть правое, то дифракционная картина в фокальной плоскости объектива, видимая глазом, не изменится. Если затем открыть оба отверстия, то одна система дифракционных колец точно совместится с другой. При этом не получится, однако, простого усиления яркости дифракционных колец. Дифракционные кольца будут пересечены параллельными интерференционными полосами, перпендикулярными к линии, соединяющей центры отверстий O_1 и O_2 (рис. 224).

Как и в опыте Юнга, эти полосы возникают из-за интерференции дифракционных пучков света, исходящих из первого и второго отверстий. Направления на соседние интерференционные максимумы (или минимумы) отличаются на угол $\vartheta = \lambda/D$. Это и есть угловое расстояние между соседними интерференционными полосами, если смотреть на фокальную плоскость из центра объектива.

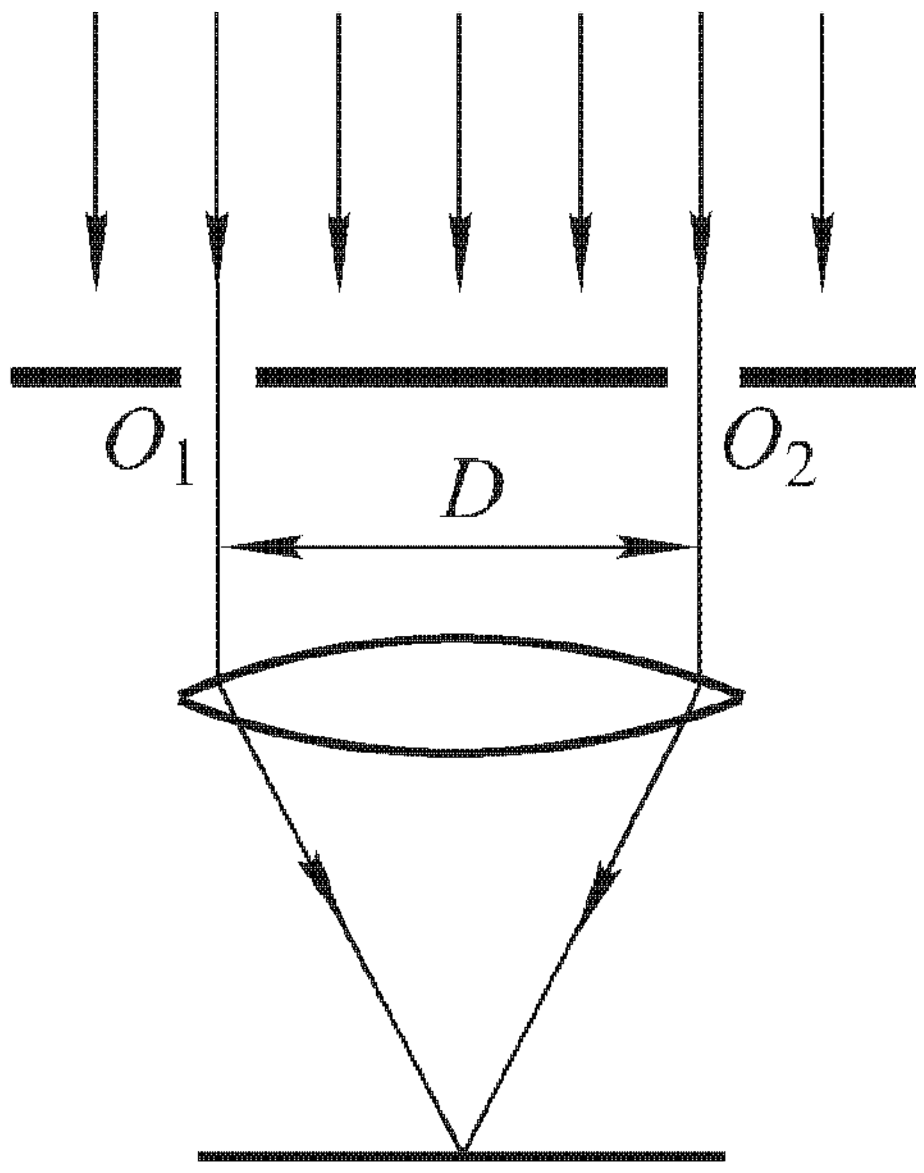


Рис. 223

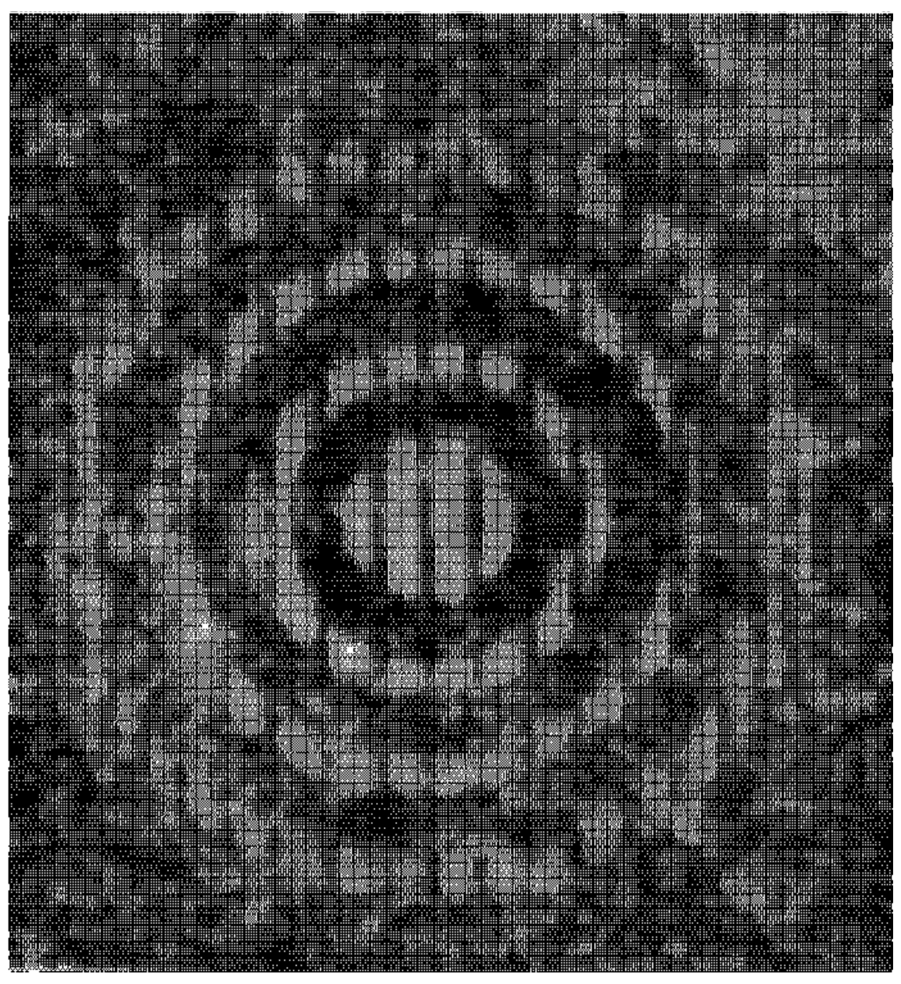


Рис. 224

Допустим, что звезда — двойная с угловым расстоянием между ее компонентами $\delta\varphi = \vartheta/2$. Тогда максимумы интерференционных полос от одной звезды наложатся на минимумы другой: интерференционные полосы либо пропадут, либо их видимость сделается наименьшей. На этом основан интерференционный метод измерения угловых расстояний между компонентами двойных звезд, идея которого была предложена Физо. Надо менять расстояние между отверстиями O_1 и O_2, пока не пропадут интерференционные полосы или их видимость сделается наименьшей. Если D — расстояние между отверстиями в этот момент, то угловое расстояние между компонентами двойной звезды найдется по формуле

$$\delta\varphi = \lambda/(2D). \tag{60.1}$$

Тот же метод применим и для измерения угловых диаметров одиночных звезд. Допустим для простоты вычислений, что звезда излучает как равномерно светящийся квадрат, плоскость которого параллельна фокальной плоскости объектива, а одна пара противоположных сторон параллельна прямой O_1O_2, соединяющей центры отверстий O_1 и O_2. Пусть угловой размер стороны квадрата будет $\delta\varphi' = \lambda/D$. Тогда можно мысленно разбить весь квадрат на пары узких одинаковых полосок, угловое расстояние между которыми равно $\lambda/(2D)$. Согласно формуле (60.1), каждая пара таких полосок не даст интерференционных полос. Следовательно, интерференционные полосы не получатся и от всей звезды. Таким образом, увеличивая расстояние между отверстиями O_1 и O_2, можно в этом случае добиться исчезновения интерференционных полос. Если D — расстояние между отверстиями в момент исчезновения, то угловой размер звезды найдется по формуле

$$\delta\varphi' = \lambda/D. \tag{60.2}$$

Если предположить, что звезда излучает как равномерно светящийся диск (что ближе соответствует действительности), то идейно простые, но громоздкие вычисления показывают, что интерференционные полосы пропадут, когда

$$\delta\varphi' = 1{,}22\frac{\lambda}{D}. \tag{60.3}$$

2. Угловые размеры $\delta\varphi'$, даваемые формулами (60.2) и (60.3), совпадают с разрешаемым расстоянием телескопа. Однако если угловой размер звезды порядка (60.2) или (60.3), то ее изображение в телескопе настолько мало отличается от изображения точечного источника, что непосредственное измерение диаметра звезд с помощью телескопа становится практически непригодным. Интерференционный же метод дает в этом случае еще хорошую точность. Однако чтобы интерференционные полосы исчезли, а это необходимо по идее самого интерференционного метода, нужен телескоп с большим диаметром объектива. Физо указал способ, как преодолеть эту трудность. Идея Физо была использована Майкельсоном, соединившим телескоп с интерферометром.

Принципиальная схема *звездного интерферометра* Майкельсона показана на рис 225. Лучи от звезды падают на два круглых отверстия O_1 и O_2 и после отражения от зеркал M_1, M_3 и M_2, M_4 попадают в объектив телескопа. Если закрыть отверстие O_2, то из-за дифракции на краях отверстия O_1 в телескоп будут видны дифракционные кольца. Если открыть O_2 и закрыть O_1, то получится такая же система колец, но сдвинутая относительно первой. Поворотом зеркала M_2 ее можно совместить с первой системой. Тогда дифракционные кольца снова будут пересечены интерференционными полосами (рис. 224). Изменяя расстояние между отверстиями O_1 и O_2 с одновременным перемещением зеркал M_1 и M_2, можно добиться либо полного исчезновения полос (в случае одиночной звезды), либо их наименьшей видимости (в случае двойной звезды, состоящей из двух различных компонентов). Зная расстояние между центрами отверстий O_1 и O_2 в этот момент, можно вычислить угловой размер звезды по формулам (60.1) или (60.3).

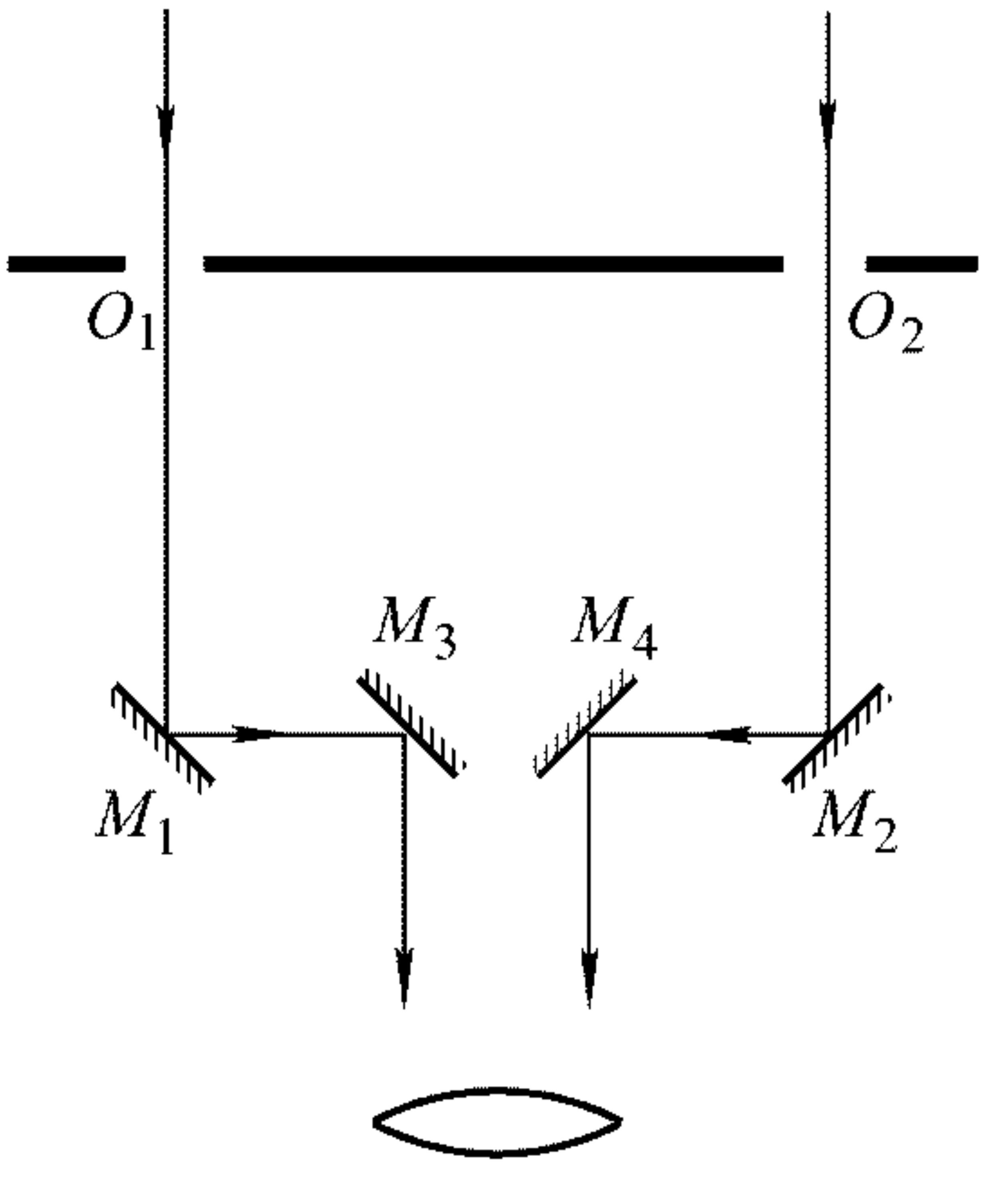


Рис. 225

3. Для уяснения действия установки Майкельсона отвлечемся от наличия звезды. Ее действие эквивалентно действию вторичных источников Гюйгенса, распределенных в плоскостях отверстий O_1 и O_2. Не теряя общности, можно для простоты рассуждений взять отверстия O_1 и O_2 бесконечно малыми. Тогда действие звезды сведется к действию вторичных точечных источников S_1 и S_2, помещенных в центрах отверстий O_1 и O_2 (рисунки 226 и 225). Если от звезды идет пучок параллельных лучей, перпендикулярных к плоскости экрана O_1O_2, то фазы вторичных источников S_1 и S_2 будут одинаковы. То же справедливо и для мнимых вторичных источников S_1', S_2' и S_1'', S_2'', являющихся изображениями S_1 и S_2 в плоских зеркалах. Таким образом, зеркала как бы сближают источники S_1 и S_2 и тем самым делают интерференционные полосы шире.

Допустим теперь, что вблизи первой звезды на угловом расстоянии $\delta\varphi$ находится вторая звезда. Волновой фронт от второй звезды будет достигать отверстий O_1 и O_2 не одновременно. Разность хода между лучами, приходящими от второй звезды в S_1 и S_2, будет $D \cdot \delta\varphi$, где D — расстояние между S_1 и S_2. Если эта разность хода равна $\lambda/2$, то вторичные источники S_1 и S_2, заменяющие действие второй звезды, будут находиться в противоположных фазах. Следовательно, максимумы интерференционных полос, даваемые второй звездой, наложатся на минимумы интерференционных полос от первой звезды. Интерференционные полосы либо пропадут, либо их видимость станет наименьшей. Это произойдет, когда $D \cdot \delta\varphi = \lambda/2$. Таким образом, мы снова приходим к формуле (60.1), причем D означает расстояние между центрами отверстий O_1 и O_2. Аналогично, в случае одиночной звезды, излучающей как равномерно светящийся квадрат или диск, получаются формулы (60.2) и (60.3).

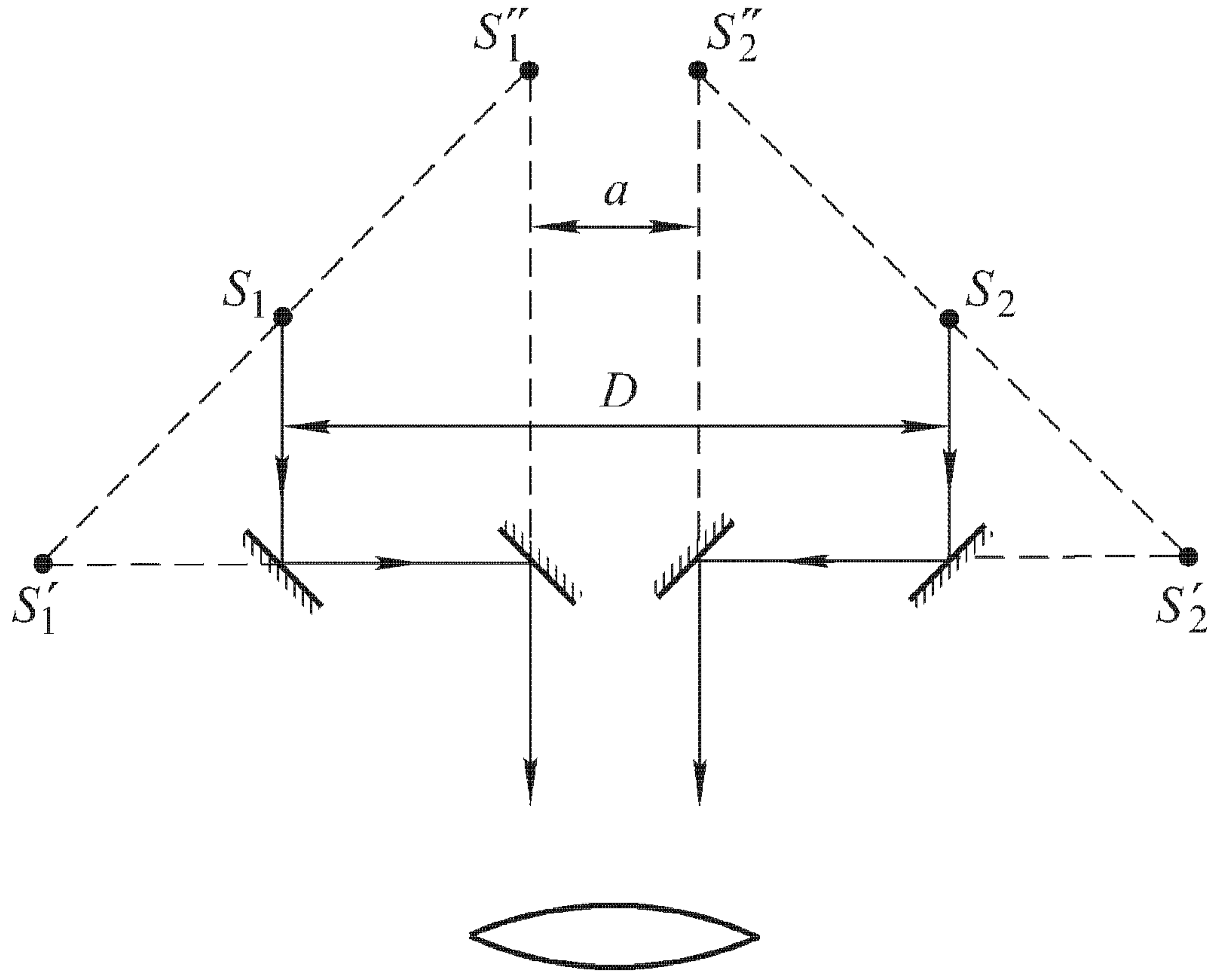


Рис. 226

Таким образом, нет надобности в телескопе с большим объективом. (Интерферометр Майкельсона был смонтирован на большом рефлекторе с диаметром зеркала 2,5 м обсерватории Маунт Вильсон, выбранном только из-за прочности своей механической конструкции. При расстоянии между зеркалами M_3 и M_4 114 см расстояние между интерференционными полосами в фокальной плоскости составляло около 0,02 мм.) Необходимо только иметь возможность достаточно далеко раздвигать отверстия O_1 и O_2 с зеркалами M_1 и M_2. Это предъявляет весьма жесткие требования к механическим качествам установки. Случайные колебания зеркал с амплитудами, составляющими незначительные доли длины волны, сделали бы измерения по этому методу невозможными. Допустим ради простоты, что центр зеркала все время остается неподвижным, а его концы испытывают беспорядочные смещения порядка h. Если бы зеркало было абсолютно твердым, то такие смещения вызвали бы беспорядочные вращения плоскости зеркала на углы порядка $2h/d$, где d — диаметр зеркала. Благодаря этому отраженные от зеркала лучи беспорядочно меняли бы свое направление на углы порядка $4h/d$. Это повело бы к дрожанию дифракционных колец. Для устойчивости интерференционных полос необходимо, чтобы угол $4h/d$ был мал по сравнению с угловым расстоянием между звездами $\lambda/(2D)$, т.е. должно быть $h \ll d\lambda/(8D)$. Хотя действительные беспорядочные колебания зеркал сложнее рассмотренных беспорядочных вращений, приведенный пример дает правильное представление о трудностях, которые должны быть преодолены при конструировании прибора. Майкельсон успешно справился с этой задачей.

4. В интерферометре Маикельсона зеркала M_1 и M_2 могли быть раздвинуты до расстояния $D = 6{,}1$ метра. Интерференционные полосы от звезды Бетельгейзе исчезали при $D = 306{,}5$ см, хотя при том же расстоянии интерференция от других звезд прекрасно наблюдалась. Считая эффективную длину волны для Бетельгейзе равной $\lambda = 5{,}75 \cdot 10^{-5}$ см, получим $\delta\varphi' = 0{,}047''$. Параллакс для этой звезды составляет $0{,}003''$. Это дает для линейного диаметра звезды около $9 \cdot 10^8$ км, что превосходит диаметр орбиты Марса.

Измерения Майкельсона были выполнены в двадцатых годах нашего столетия, когда астрофизика в современном ее понимании еще только зарождалась. Результаты этих измерений произвели сильное впечатление на современников. Майкельсон начал строить интерферометр с базой $D = 15{,}24$ м, но смерть помешала ему осуществить это намерение.

Звездным интерферометром Майкельсона были измерены угловые диаметры только нескольких звезд. Все они, как и Бетельгейзе, — гиганты, линейные диаметры которых во много раз превосходят диаметр Солнца. Одна из причин небольшого числа измеренных диаметров звезд связана с вредным влиянием турбулентности атмосферы. Правда, это влияние значительно сильнее проявляется при наблюдении в телескоп, чем при работе с интерферометром. В случае интерферометра изменение показателя преломления воздуха перед небольшими отверстиями зеркал лишь смещает интерференционную картину в целом, так что интерференционные полосы остаются различимыми, если только это смещение происходит достаточно медленно. При наблюдении же в тех же условиях в телескоп с недиафрагмированным объективом нерегулярные изменения показателя преломления воздуха сильно искажают изображение звезды.

Другая причина заключается в том, что линейные диаметры подавляющего большинства звезд мало отличаются от диаметра Солнца. На расстоянии же ближайшей звезды диск Солнца был бы виден под углом всего $0{,}007''$. Для измерения таких углов расстояние между внешними зеркалами интерферометра M_1 и M_2 (база) должно быть не меньше примерно 20 м. Постройка такого большого интерферометра трудна из-за высоких требований к механическим параметрам конструкции всей установки, о которых говорилось выше.

5. Принципы, изложенные выше, применимы, конечно, и в радиодиапазоне. Радиотелескопы, благодаря большим значениям длин радиоволн, характеризуются низкой разрешающей способностью. Даже для самых крупных из них с диаметром зеркала $D \sim 100$ м, работающих на длине волны $\lambda \sim 10$ см, разрешаемое угловое расстояние $\sim \lambda/D \approx 10^{-3}$ рад, т.е. всего около $3'$. Иное дело — радиоинтерферометр, работающий по принципу звездного интерферометра Майкельсона. Он состоит из двух антенн, установленных в разных местах, сигналы с которых подаются на общий детектор. Высокая разрешающая способность достигается за счет увеличения расстояния D между антеннами (базы радиоинтерферометра). Рекордное угловое разрешение достигается, когда антенны удалены друг от друга на межконтинентальные расстояния. Были осуществлены радиоинтерферометры с базой Крым–США, США–Австралия. На самой короткой длине волны было достигнуто угловое разрешение $\sim 10^{-4}$ угловой секунды, т.е. разрешающая способность была примерно в 100 раз больше, чем у звездного интерферометра Майкельсона.

Надо отметить, что с увеличением базы возрастают технические трудности, обусловленные внесением нерегулярно меняющейся разности фаз на пути от антенн к детектору. Для устранения этого недостатка в радиоинтерферометре Брауна и Твисса сигналы, принятые антеннами, детектируются независимо. Измеряется корреляция флуктуирующих интенсивностей этих сигналов в зависимости от расстояния между антеннами. По этим данным находятся угловые размеры источника. Браун и Твисс показали, что аналогичное устройство может работать также и в видимой области спектра. Свет от звезды фокусируется двумя вогнутыми зеркалами на два фотоэлемента. Измеряется корреляция флуктуирующих фототоков, возбуждающихся в фотоэлементах, в зависимости от расстояния между зеркалами. Без особых осложнений расстояние между зеркалами может быть сделано большим и достигнута высокая разрешающая способность прибора.

§ 61. Дифракция на двумерных и трехмерных решетках. Дифракция рентгеновских лучей

1. *Двумерной решеткой* называется всякая структура, свойства которой периодически меняются в двух различных направлениях. Примером могут служить две скрещенные одномерные решетки, т.е. решетки, наложенные одна на другую таким образом, что штрихи одной решетки пересекаются со штрихами другой под некоторым углом. Дифракционная картина от такой структуры, как нетрудно показать, может быть получена путем наложения дифракционных картин от соответствующих одномерных решеток.

Трехмерные, пространственные решетки обладают периодичностью в трех различных направлениях. Они играют важную роль в физике рентгеновских лучей. Долгое время не удавалось получить дифракцию рентгеновских лучей, так как дифракционные решетки, применявшиеся в оптической области спектра, были слишком грубыми для рентгеновских лучей с длиной волны порядка 0,1 нм и меньше (см., впрочем, § 46, п. 8). Лауэ (1879–1960) в 1912 г. предложил для этой цели воспользоваться кристаллом, который можно рассматривать как *естественную периодическую пространственную решетку*, построенную из атомов или ионов. По указанию Лауэ его сотрудники Фридрих (1883–1968) и Книппинг (1883–1935) направили узкий пучок сплошного рентгеновского излучения на монокристалл и получили на фотопластинке картину дискретных, правильно расположенных пятен, возникших в результате дифракции рентгеновских лучей на кристаллической решетке. Этот основополагающий опыт установил *волновую природу* рентгеновского излучения и заложил начала современной *физики твердого тела.* Он впервые принес экспериментальное доказательство давно высказывавшейся кристаллографами гипотезы о структуре кристаллов из частиц, правильно и периодически расположенных в пространстве.

Из этого опыта возникли два направления в физике: *рентгеновская спектроскопия* и *рентгеноструктурный анализ.* Рентгеновская спектроскопия использует естественные кристаллы известной кристаллической структуры для анализа рентгеновского излучения и измерения длин волн. Рентгеноструктурный анализ, напротив, использует рентгеновское излучение известной длины волны для выяснения кристаллической структуры кристаллов и измерения параметров этой структуры. По сравнению с оптической дифракцией дифракция рентгеновских лучей обладает той особенностью, что она осуществляется *без линз и зеркал*, ввиду отсутствия материалов, из которых можно было бы изготовить эти приспособления.

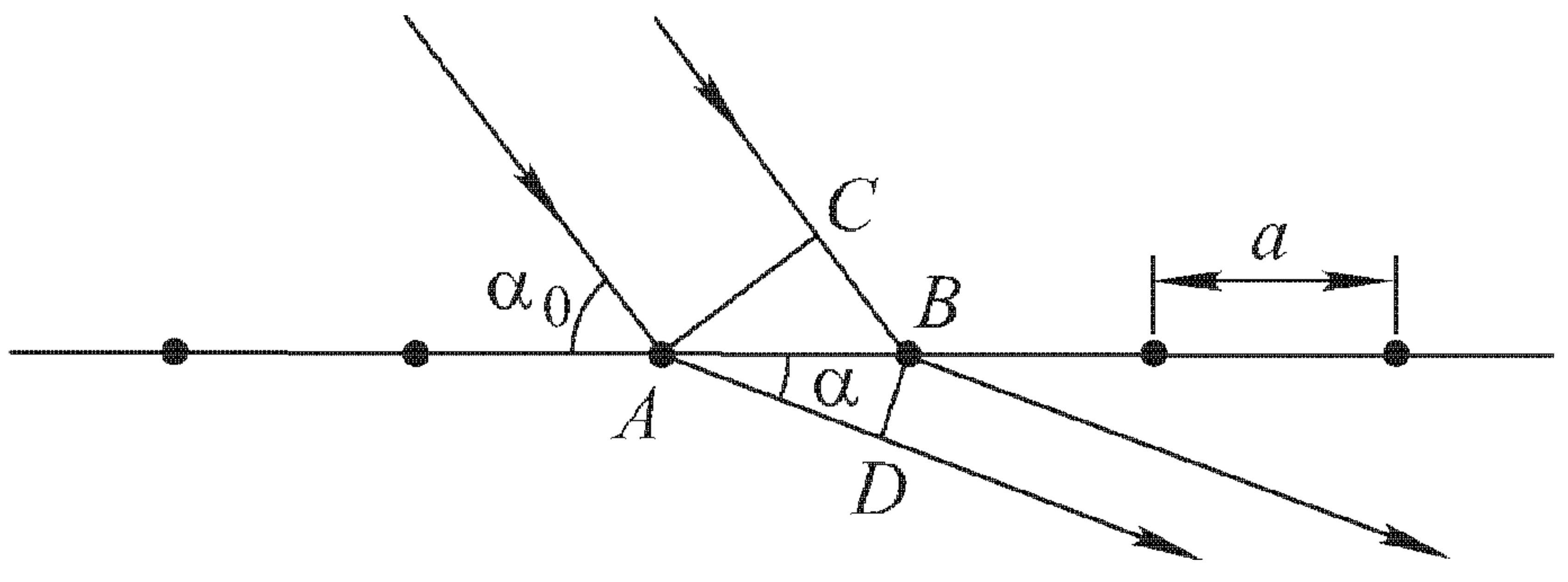


Рис. 227

2. Качественное рассмотрение дифракции рентгеновских лучей удобно начать с *дифракции на прямолинейной цепочке*, состоящей из одинаковых равноотстоящих частиц (атомов). Расстояние между соседними частицами обозначим через a. Пусть на такую цепочку под углом скольжения α_0 падает параллельный пучок рентгеновских лучей (рис. 227). Разность хода между лучами, рассеянными соседними атомами под углом α, будет $AD - CB = a(\cos\alpha - \cos\alpha_0)$. Условие интерференционного усиления этих лучей имеет вид

$$a(\cos\alpha - \cos\alpha_0) = m\lambda \tag{61.1}$$

($m = 0, \pm 1, \pm 2, \ldots$). Оно определяет *положения дифракционных максимумов в фраунгоферовой дифракционной картине*, т.е. на бесконечных расстояниях от цепочки. Бесконечными могут считаться расстояния r, удовлетворяющие условию:

$$r \gg l^2/\lambda, \tag{61.2}$$

где l — длина цепочки. Ввиду малости длин рентгеновских волн, эти расстояния всегда очень велики. Так, при $l = 1$ мм, $\lambda = 0{,}1$ нм $= 10^{-7}$ мм условие (61.2) дает $r \gg 10^7$ мм $= 10$ км. Таким образом, для осуществления фраунгоферовой дифракции потребовались бы очень большие расстояния. В реальных опытах фотографическая пластинка, регистрирующая дифракционную картину, ставится на расстоянии в несколько десятков сантиметров от рассеивающего кристалла или меньше, т.е. не в волновой зоне, а в десятки и сотни тысяч раз ближе. Это замечание относится, конечно, не только к линейным цепочкам, но и к поверхностным и объемным решеткам.

Однако, как было выяснено в § 52, *волновое поле, возникающее при дифракции на плоской решетке, представляется на всех расстояниях суперпозицией плоских волн различных направлений.* В эту суперпозицию входят и *неоднородные волны.* Но на расстояниях от решетки, превышающих ее период, неоднородные волны не играют роли, так как на таких расстояниях они практически полностью затухают. Так же обстоит дело и в случае дифракции на кристаллической решетке. (В случае линейной цепочки вместо плоских удобнее пользоваться *цилиндрическими волнами.*) Фотопластинка ставится на малых расстояниях от кристалла, где применима геометрическая оптика, но все же достаточно далеко, чтобы различные плоские волны, исходящие от кристалла, успели разделиться пространственно. Для определения положения максимумов интенсивности на фотопластинке достаточно знать *направления* распространения этих волн. Но эти направления не зависят от расстояния до кристалла. В области применимости геометрической оптики они такие же, что и в волновой зоне, а потому для определения положения максимумов интенсивности на фотопластинке можно пользоваться формулой (61.1). Однако такие максимумы отличаются от интерференционных максимумов в волновой зоне. В каждый максимум в волновой зоне колебания от всех атомов решетки приходят *либо в одинаковых фазах, либо в фазах, отличающихся на* $2m\pi$ (m — целые числа). Для максимумов же интенсивности в области применимости геометрической оптики это не имеет места.

При фиксированном угле α_0 условие (61.1) определяет дискретный набор углов α, удовлетворяющих этому условию. Оно выделяет в пространстве *дискретное семейство конусов*, вдоль образующих которых могут распространяться дифрагированные пучки лучей. В сечении таких конусов плоскостью фотопластинки получается *дискретное семейство эллипсов или гипербол* в зависимости от направления этой плоскости. В частности, когда плоскость пластинки перпендикулярна к направлению цепочки, возникает *семейство концентрических кругов.*

3. Двумерные и трехмерные решетки могут быть *простыми* (*примитивными*) и *составными* (см. т. II, § 130). Решетка называется простой, если она построена из одинаковых атомов, причем элементарная ячейка решетки состоит из восьми атомов, расположенных в вершинах параллелепипедов. Все остальные решетки называются составными. Составная решетка состоит из нескольких простых решеток, вставленных друг в друга. Дифракционная картина, возникающая при дифракции рентгеновских волн на составной решетке, получается в результате интерференции дифракционных картин от простых решеток, из которых она состоит. Поэтому достаточно ограничиться рассмотрением дифракции на простых решетках.

Если какие-либо два атома простой решетки соединить прямой линией, то, ввиду периодичности распределения атомов в пространстве, на этой прямой окажется бесконечно много атомов, находящихся на одинаковых расстояниях друг от друга. Такие прямые мы будем называть *атомными прямыми*, а плоскости, в которых располагаются атомы, — *атомными плоскостями*. Всю неограниченную решетку можно рассматривать как бесконечную двоякопериодическую систему параллельных атомных прямых или как бесконечную однократно периодическую систему параллельных атомных плоскостей. Оба эти представления не единственны, а могут быть выполнены бесконечным множеством способов. Три произвольные атомные прямые, не лежащие в одной плоскости и пересекающиеся в каком-либо атоме, можно принять за координатные оси X, Y, Z прямолинейной (вообще говоря, косоугольной) системы координат. Тогда координаты атомов простой решетки представятся выражениями

$$\begin{gathered} x_{lmn} = la_1, \quad y_{lmn} = ma_2, \quad z_{lmn} = na_3 \\ (l, m, n = 0, \pm 1, \pm, 2, \ldots), \end{gathered} \tag{61.3}$$

где a_1, a_2, a_3 — постоянные, называемые *периодами решетки*. Элементарной ячейкой такой решетки является параллелепипед с ребрами a_1, a_2, a_3, в вершинах которого находятся атомы.

Пусть на простую решетку падает параллельный пучок рентгеновских лучей, образующий углы α_0, β_0, γ_0 с координатными осями X, Y, Z. Чтобы волны, рассеянные всеми атомами в направлении прямой, составляющей углы α, β, γ с координатными осями, при интерференции в волновой зоне усиливали друг друга, должны выполняться условия

$$\begin{gathered} a_1(\cos\alpha - \cos\alpha_0) = m_1\lambda, \\ a_2(\cos\beta - \cos\beta_0) = m_2\lambda, \\ a_3(\cos\gamma - \cos\gamma_0) = m_3\lambda \\ (m_1, m_2, m_3 = 0, \pm 1, \pm 2, \ldots), \end{gathered} \tag{61.4}$$

называемые *условиями Лауэ*. Необходимость первого условия станет очевидной, если заметить, что оно является условием интерференционного усиления волн, рассеянных под углом α к оси X атомами каждой атомной прямой, параллельной этой оси. Аналогичный смысл имеют и остальные два условия.

Но условия (61.4) являются и достаточными для интерференционного усиления волн, рассеянных в рассматриваемом направлении всеми атомами решетки. Действительно, проведем через произвольный атом 1 атомную прямую, параллельную оси X. При выполнении первого условия (61.4) в направлении под углом α к этой прямой получится интерференционный максимум. Проведем теперь через тот же атом 1 атомную прямую, параллельную оси Y. При выполнении второго условия (61.4) все атомы этой прямой рассеивают волны в рассматриваемом направлении в той же фазе, что и атом 1. Значит, все атомы обеих атомных прямых, а с ними и все атомы, лежащие в их плоскости, будут посылать волны в том же направлении также в одинаковых фазах. Таким образом, выполнение первых двух условий (61.4) приводит к интерференционному усилению волн, рассеиваемых в рассматриваемом направлении всеми атомами любой атомной плоскости, параллельной координатной плоскости XY. Аналогично убедимся, что при выполнении еще третьего условия (61.4) будет иметь место интерференционное усиление волн, рассеянных всеми такими атомными плоскостями. Тем самым достаточность условий (61.4) доказана.

4. Дифракционная картина, возникающая на фотопластинке, поставленной на пути рентгеновских пучков, рассеянных монокристаллом в опытах типа Лауэ, называется *лауэграммой.* Об использовании условий Лауэ в области применимости геометрической оптики можно повторить все, что было сказано выше в связи с формулой (61.1). *Формулы Лауэ* (61.4) *указывают направления пучков, возникающих при дифракции на кристалле.* Физический смысл лауэграммы хорошо иллюстрируется аналогией с отражением светового пучка от многогранного зеркала. Здесь возникают отраженные пучки, распространяющиеся в различных направлениях. При падении на экран они дают систему правильно расположенных светлых пятен, аналогичную лауэграмме, возникающей при дифракции рентгеновских лучей.

Рассмотрим сначала лауэграмму от плоской двумерной кристаллической решетки. В этом случае три условия Лауэ (61.4) сводятся к двум. Если плоскость решетки принять за атомную плоскость XY, то останутся только два первых условия (61.4). Первое условие (61.4) означает, что максимумы лежат на поверхности конуса, образующего угол α с осью X, а второе — на поверхности другого конуса, образующего угол β с осью Y. Прямые, по которым пересекаются поверхности обоих конусов, указывают направления дифрагированных пучков. При пересечении таких пучков плоскостью фотопластинки возникают дискретные дифракционные точечные максимумы, расположенные вдоль эллипсов, гипербол или кругов в зависимости от направления этой плоскости. Максимумы могут и не получиться, если

$\cos\alpha$ или $\cos\beta$, или обе эти величины, вычисленные по формуле (61.4), окажутся по модулю больше единицы.

Рассмотрим теперь трехмерную решетку. К двум конусам, выделяемым первыми двумя условиями (61.4), теперь добавляется еще третий конус, образующие которого составляют угол γ с осью Z. Дифрагированные пучки должны одновременно лежать на поверхности всех трех конусов. Но три конуса, вообще говоря, не пересекаются вдоль общей прямой. Отсюда следует, что при падении на монокристалл монохроматического рентгеновского луча дискретные пучки рассеянного излучения, вообще говоря, не возникают, а рассеяние происходит более или менее равномерно во все стороны. Исключение составляет только прямой луч, проходящий через кристалл без изменения направления. Но для избранных длин волн три конуса могут иметь общие образующие. Поэтому *для получения лауэграмм от трехмерных решеток необходимо сплошное рентгеновское излучение*, так как в таком излучении могут присутствовать такие длины волн, для которых выполняются все три условия Лауэ (61.4). Таким образом, если бы глаз обладал способностью воспринимать рентгеновские лучи и различать их цвета, то лауэграмма, наблюдаемая на экране, представлялась бы для него цветной, т.е. состоящей из пятен разного цвета.

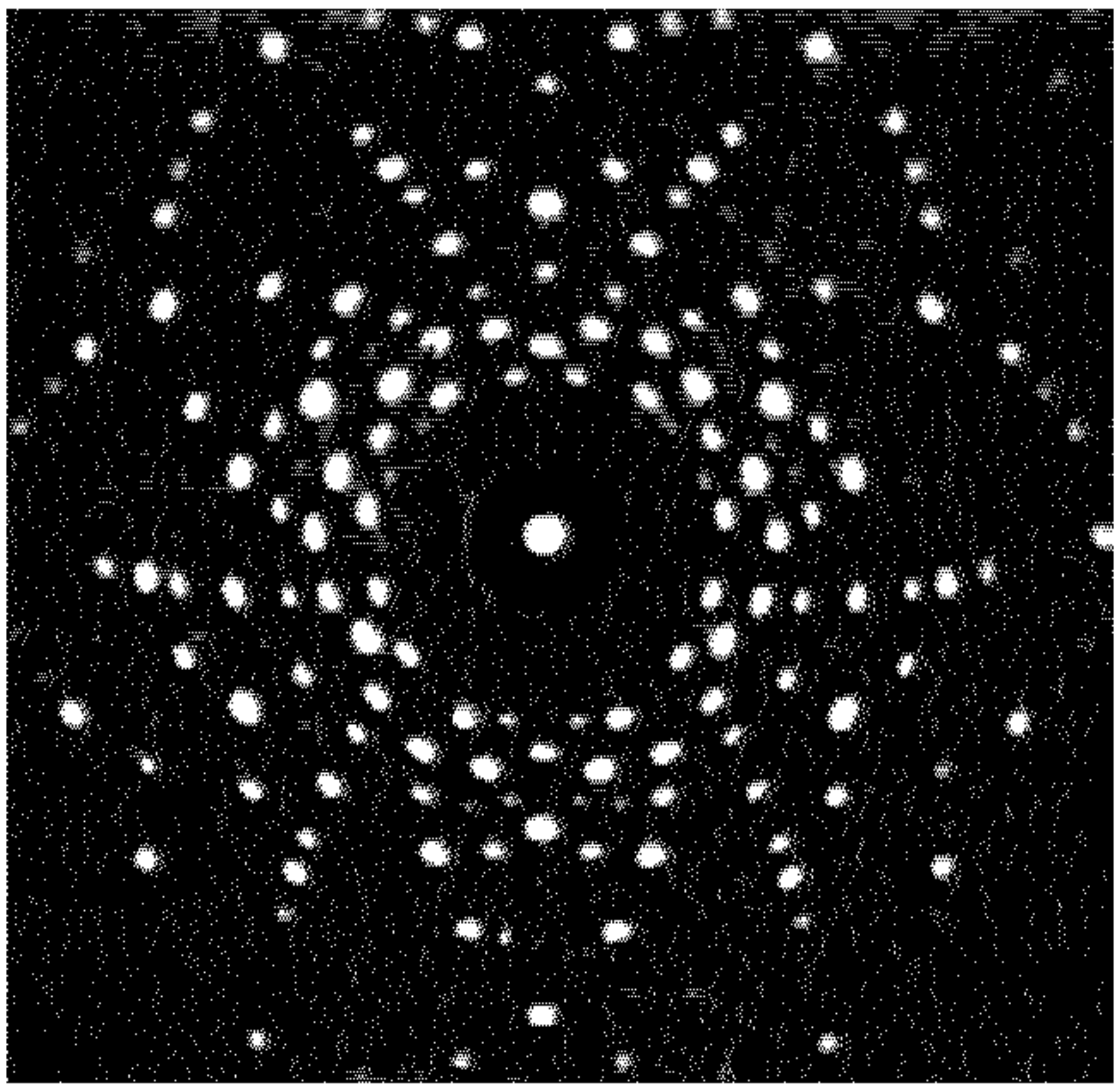


Рис. 228

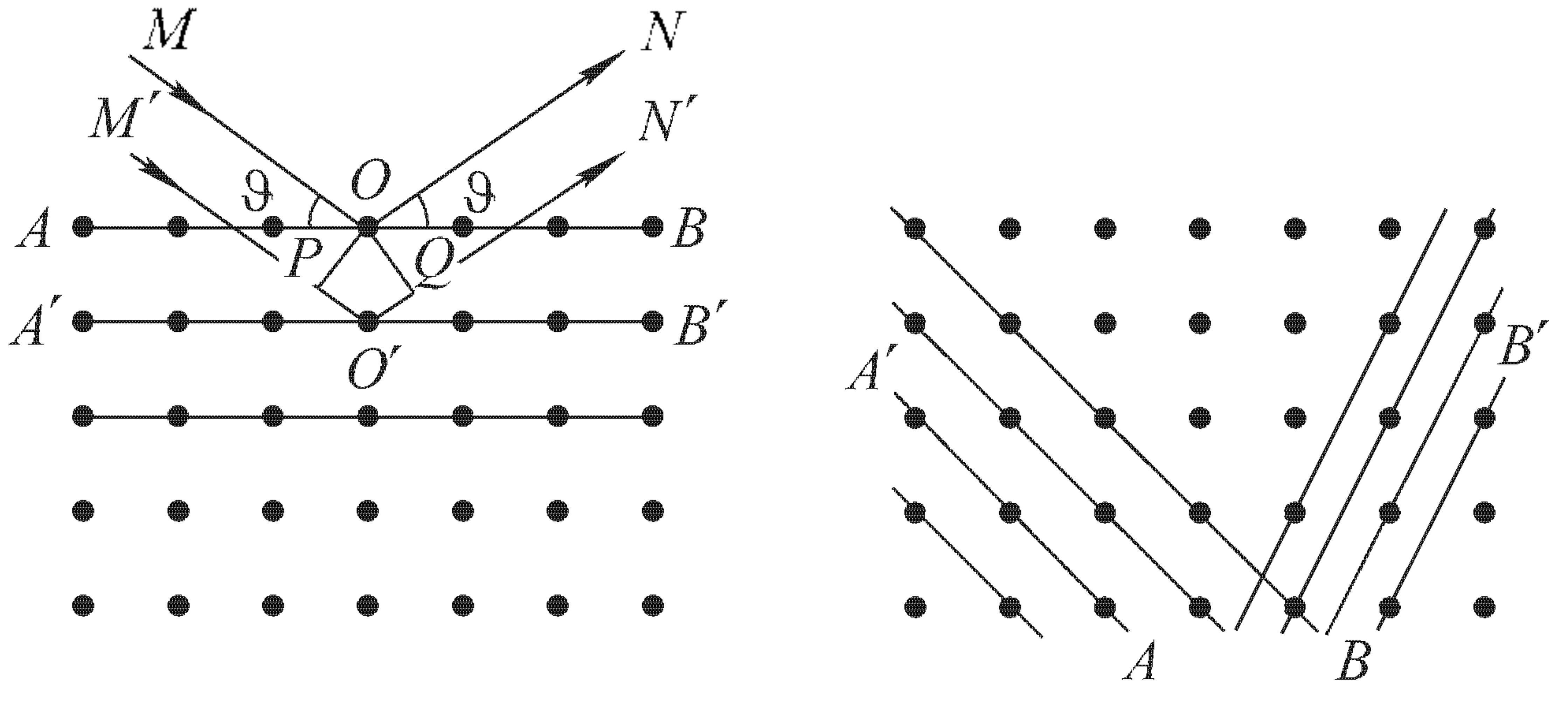
Подтвердим полученные результаты простым вычислением. Предположим, что простая кристаллическая решетка принадлежит к ромбической системе, так что элементарная ячейка будет прямоугольным параллелепипедом. Пусть для некоторой длины волны λ условия Лауэ (61.4) выполняются. Разрешим их относительно $\cos\alpha$, $\cos\beta$, $\cos\gamma$, возведем в квадрат и сложим почленно. Тогда, принимая во внимание соотношения $\cos^2\alpha + {}$ $+\cos^2\beta+\cos^2\gamma = 1$, $\cos^2\alpha_0+\cos^2\beta_0+\cos^2\gamma_0 = 1$, после простых преобразований получим

$$\lambda = -\frac{2(m_1\cos\alpha_0 + m_2\cos\beta_0 + m_3\cos\gamma_0)}{m_1^2 + m_2^2 + m_3^2}. \tag{61.5}$$

Только при этом условии уравнения (61.4) совместны. Оно однозначно определяет длину волны, при которой может получиться дифракционный максимум, выделяемый целыми числами m_1, m_2, m_3 и углами α, β, γ.

На рис. 228 приведена лауэграмма для монокристалла кварца. Закономерное расположение пятен на фотопластинке указывает на правильность расположения атомов, из которых построен кристалл.

5. Дифракцию рентгеновских лучей в кристаллах можно трактовать несколько иначе. Такая трактовка была дана английским физиком Лоуренсом Брэггом (1890–1971) и независимо от него русским кристаллографом Ю.В. Вульфом (1863–1925). Проведем в кристалле произвольную атомную плоскость AB, достаточно густо усеянную атомами (рис. 229 *а*). Если на нее падает рентгеновский луч MO, то под тем же углом возникнет отраженный луч ON. В том же направлении возникнут лучи, отраженные атомными плоскостями, параллельными плоскости AB. Интенсивность луча, отраженного отдельной атомной плоскостью, слишком мала, чтобы произвести заметное действие. Последнее может возникнуть лишь в результате интерференционного усиления всех лучей, отраженных рассматриваемыми атомными плоскостями. Разность хода $PO'Q$ между лучами, отраженными соседними плоскостями, равна $2d\sin\vartheta$, где d — расстояние между этими плоскостями. Для интерференционного усиления должно выполняться условие



а *б*

Рис. 229

$$2d\sin\vartheta = m\lambda, \tag{61.6}$$

называемое *условием Брэгга–Вульфа* ($m = 0, \pm1, \pm2, \ldots$). Каков бы ни был угол скольжения ϑ и порядок отражения m, всегда найдутся длины волн λ, удовлетворяющие условию (61.6). Только волны с такими длинами могут отражаться от рассматриваемых атомных плоскостей.

В кристалле можно провести бесконечное множество систем параллельных атомных плоскостей в различных направлениях. Таковы, например, плоскости, параллельные атомной плоскости AA' или атомной плоскости BB' (рис. 229 *б*). Эффективными являются только такие плоскости, на которых атомы расположены достаточно часто. От всех этих плоскостей возможно интерференционное отражение. И дифракционную картину можно рассматривать как совокупность рентгеновских пучков, претерпевших отражения на таких атомных плоскостях.

Против последнего заключения можно, однако, выдвинуть следующее возражение. Рентгеновский пучок MO, падающий на атомную плоскость AB (рис. 229 *а*), дает не только отраженный пучок ON, но и боковые дифрагированные пучки, которые при определенных условиях могут усиливаться пучками того же направления, дифрагировавшими на параллельных атомных плоскостях. Такие пучки в наших рассуждениях не были учтены. Поэтому может показаться, что метод Брэгга–Вульфа дает не все возможные дифракционные пучки в дифракционной картине. Следующее простое рассуждение, устанавливающее *эквивалентность условий Лауэ и условия Брэгга–Вульфа*, показывает, что это не так.

6. Примем векторы $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, являющиеся ребрами элементарного параллелепипеда кристаллической решетки, за базисные векторы косоугольной системы координат. Тогда радиус-вектор каждого атома решетки представится выражением

$$\mathbf{r} = x\mathbf{a}_1 + y\mathbf{a}_2 + z\mathbf{a}_3, \qquad (61.7)$$

в котором координаты x, y, z принимают целочисленные значения. Пусть $\mathbf{s}_0$ — единичный вектор, проведенный в направлении падающего луча, а $\mathbf{s}$ — единичный вектор, указывающий направление одного из дифрагированных пучков (рис. 230). Тогда условия Лауэ (61.4) можно записать в следующей векторной форме:

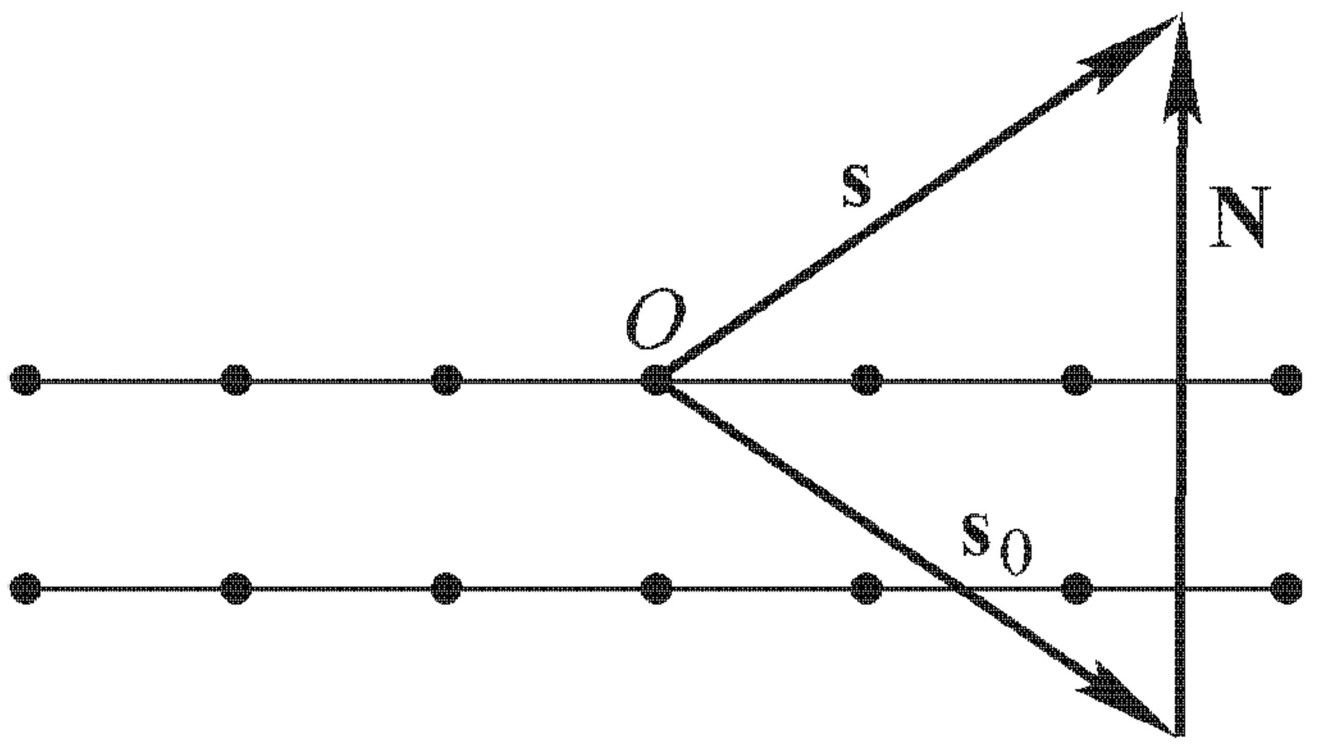


Рис. 230

$$\begin{aligned}(\mathbf{s}-\mathbf{s}_0)\mathbf{a}_1 &= m_1\lambda,\\ (\mathbf{s}-\mathbf{s}_0)\mathbf{a}_2 &= m_2\lambda, \qquad (61.8)\\ (\mathbf{s}-\mathbf{s}_0)\mathbf{a}_3 &= m_3\lambda.\end{aligned}$$

Вектор $\mathbf{N} = \mathbf{s} - \mathbf{s}_0$ направлен параллельно биссектрисе угла, образованного падающим и отраженным лучами. Введя этот вектор, получим

$$(\mathbf{Na}_1) = m_1\lambda, \quad (\mathbf{Na}_2) = m_2\lambda, \quad (\mathbf{Na}_3) = m_3\lambda. \qquad (61.9)$$

Через атом, находящийся в начале координат O, проведем плоскость, перпендикулярную к вектору $\mathbf{N}$. Докажем, что она является атомной плоскостью. Уравнение рассматриваемой плоскости имеет вид $(\mathbf{Nr}) = 0$. Чтобы атом с координатами (61.7) лежал в этой плоскости, необходимо и достаточно, чтобы его координаты удовлетворяли уравнению

$$(\mathbf{Na}_1)x + (\mathbf{Na}_2)y + (\mathbf{Na}_3)z = 0,$$

или, ввиду соотношений (61.9), $m_1x + m_2y + m_3z = 0$. Каковы бы ни были целые числа m_1, m_2, m_3, существует двухпараметрическое семейство целочисленных решений этого уравнения. Тем самым доказано, что плоскость $(\mathbf{Nr}) = 0$ является атомной плоскостью.

Из доказанного следует, что для любого дифрагированного луча $\mathbf{s}$ можно указать атомную плоскость, а следовательно, и бесконечное семейство параллельных ей атомных плоскостей, при зеркальном отражении от которых возникают лучи того же направления, что и рассматриваемый дифрагированный луч. Тем самым доказано, что условием Брэгга–Вульфа охватываются все направления, по которым могут распространяться дифрагированные рентгеновские пучки. Значит, *каждый боковой дифракционный пучок, возникший при дифракции на той или иной атомной плоскости, совпадает по направлению с пучком, зеркально отразившимся какой-то другой атомной плоскостью. Направлениями зеркально отраженных лучей исчерпываются все возможные направления на дифракционные максимумы.*

Конечно, не всякие атомные плоскости эффективно отражают и дают максимумы, действительно наблюдающиеся на опыте. Необходимо, чтобы атомные плоскости были усеяны атомами *достаточно густо.* Иначе интенсивность отраженных лучей может оказаться настолько малой, что они не проявят никакого действия на опыте.

7. В связи с изложенным уточним смысл условия Брэгга–Вульфа. Выделим какое-либо семейство параллельных атомных плоскостей и рассмотрим лучи, возникшие при зеркальном отражении от каждой из этих плоскостей в отсутствие остальных. Условие Брэгга–Вульфа вовсе не означает, что при интерференции таких лучей между собой и с падающим лучом возникнет истинная картина распределения волнового поля в кристалле. Действительно, луч, который при выводе и интерпретации условия Брэгга–Вульфа принято называть лучом, отраженным отдельной атомной плоскостью, в действительности не является таковым. Он возникает в результате сложного процесса, в котором участвуют *атомы всего кристалла,* а не только атомы рассматриваемой атомной плоскости. В частности, в формировании этого луча участвуют боковые пучки того же направления, возникающие при дифракции на других атомных плоскостях. Однако окончательная дифракционная картина будет такой, как если бы отдельные атомные плоскости только зеркально отражали рентгеновские лучи с некоторыми надлежаще выбранными эффективными коэффициентами отражения и не давали никаких боковых дифракционных пучков.

8. Изложенным по существу завершается доказательство эквивалентности методов Лауэ и Брэгга–Вульфа. Не лишено, однако, интереса показать, что из формул Лауэ без привлечения каких бы то ни было дополнительных соображений вытекает условие Брэгга–Вульфа. Вычислим прежде всего длину вектора $\mathbf{N} = \mathbf{s} - \mathbf{s}_0$, нормального к плоскости, от которой происходит рассматриваемое брэгговское отражение. Угол между единичными векторами $\mathbf{s}_0$ и $\mathbf{s}$ равен 2ϑ, т.е. удвоенному углу скольжения. Следовательно, $\mathbf{N}^2 = 2 - 2(\mathbf{s}\mathbf{s}_0) = 2(1 - \cos 2\vartheta) = 4\sin^2\vartheta$, а потому $N = 2\sin\vartheta$. Примем атомную плоскость, от которой происходит рассматриваемое брэгговское отражение, за координатную плоскость XY. Тогда межплоскостное расстояние d представится выражением

$$d = \left(\mathbf{a}_3\frac{\mathbf{N}}{N}\right) = \frac{(\mathbf{a}_3\mathbf{N})}{2\sin\vartheta}.$$

Используя последнее соотношение (61.9), получим

$$d = m_3\lambda/(2\sin\vartheta),$$

что совпадает с условием Брэгга–Вульфа (61.6).

9. Рассмотрим два примера на применение условия (61.6).

а) Рентгеновский спектрограф с вращающимся кристаллом (рис. 231). Рентгеновский пучок после диафрагмирования на щелях направляется на монокристалл K известной кристаллической структуры. Кристалл может вращаться вокруг оси, перпендикулярной к падающему лучу, или совершать вращательные колебания вокруг этой оси. Дифрагированный пучок попадает на фотопластинку FF' или в ионизационную камеру. Таким образом, направления падающего и дифрагированного пучков, а с ними и величина угла скольжения ϑ фиксированы. Если падающий пучок монохроматичен или состоит из отдельных монохроматических линий, то при произвольном положении кристалла условие Брэгга–Вульфа, вообще говоря, не будет выполняться. Однако при повороте кристалл может занять такое положение, при котором это условие выполняется. Тогда возникнет отраженный максимум, регистрируемый по почернению фотопластинки или по току в ионизационной камере. Найдя такое положение, можно определить угол скольжения ϑ, а затем по формуле (61.6) вычислить длину волны λ.

Разумеется, такой метод дает не абсолютное значение длины волны, а только ее отношение к постоянной решетки d. Для абсолютных измерений надо знать величину d. Для нахождения d достаточно независимым способом измерить длину волны какой-либо строго определенной спектральной линии. Это можно сделать, например, с помощью обычной отражательной дифракционной решетки с известным периодом, измерив угол дифракции при скользящем падении луча (см. § 46, п. 8).

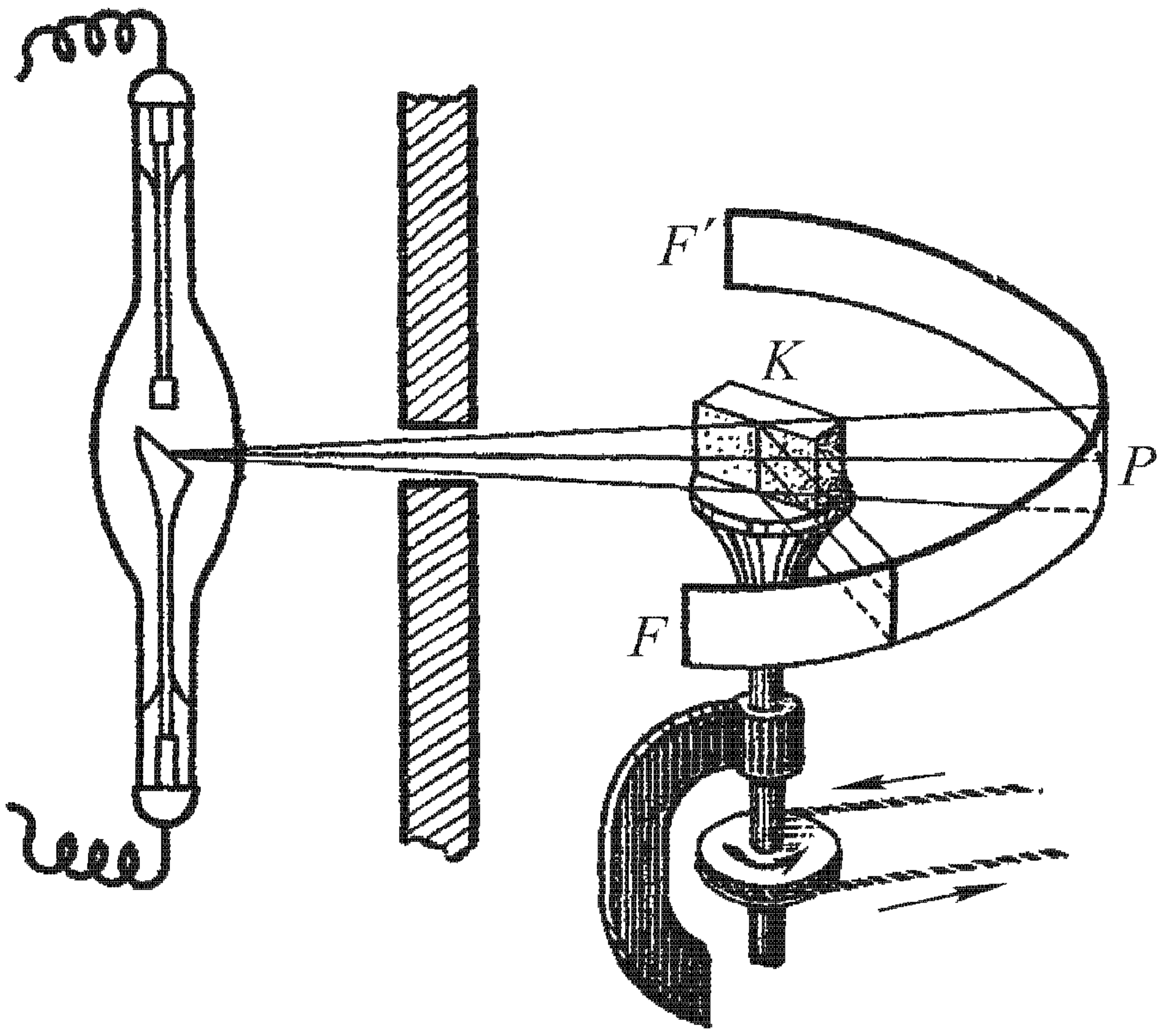


Рис. 231

б) Метод Дебая–Шерера–Хелла. Этот метод применяется в рентгеноструктурном анализе для исследования кристаллической структуры металлов и других кристаллических материалов в порошкообразном состоянии. Исследуемый образец (поликристалл) обычно имеет форму цилиндрика и состоит из множества мельчайших кристалликов, беспорядочно ориентированных во всевозможных направлениях. Установка аналогична приведенной на рис. 231. Однако столик, на котором помещается исследуемый образец, неподвижен. На образец направляется монохроматический рентгеновский луч с известной длиной волны λ. Дифракционная картина, называемая *дебаеграммой*, фотографируется.

Происхождение этой картины объясняется следующим образом. Среди множества беспорядочно ориентированных кристалликов найдется еще очень много кристалликов с такими ориентациями, что при заданной длине волны λ будет выполнено условие Брэгга–Вульфа. Лучи, испытавшие брэгговские отражения от таких кристалликов, образуют поверхность конуса, ось которого направлена вдоль падающего луча, а угол раствора определяется межплоскостным расстоянием d (см. рис. 232, на котором показано отражение от отдельного микрокристаллика; сам микрокристаллик изображен в виде зеркальца). Так как эти расстояния образуют дискретный набор, то за образцом возникнет дискретное семейство конусов с общими вершиной и осью.

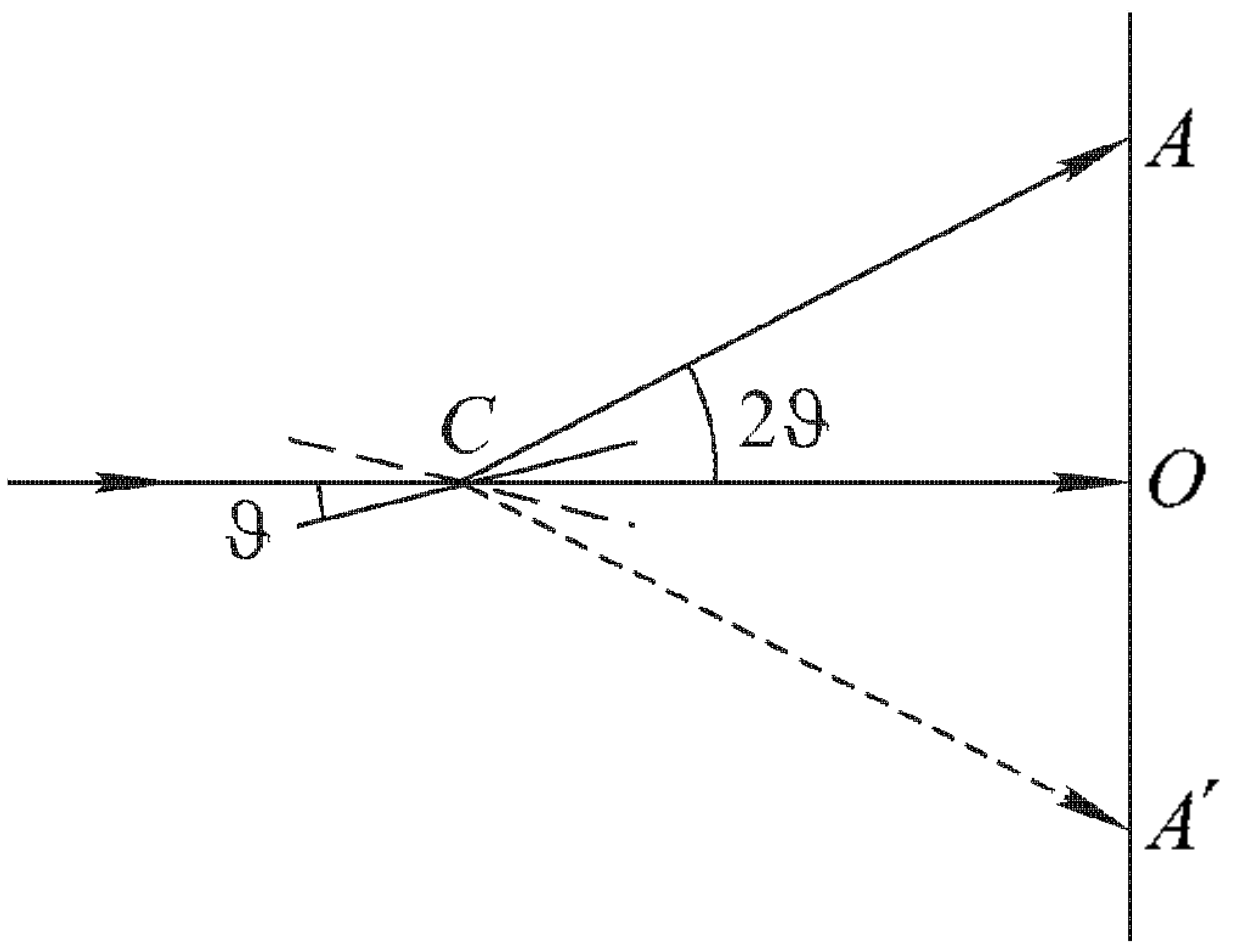


Рис. 232

Если бы фотопластинка была установлена перпендикулярно к этой общей оси, то дебаеграмма состояла бы из концентрических кругов. Измерив радиусы этих кругов, можно определить возможные значения угла ϑ, а затем по формуле Брэгга–Вульфа вычислить соответствующие межплоскостные расстояния и воспользоваться этими данными для воспроизведения кристаллической структуры образца. Чтобы получить действительно все межплоскостные расстояния, фотопластинке придают форму полоски, опоясывающей по окружности исследуемый образец. Дебаеграмма, полученная на кристаллическом порошке NaCl, приведена на рис. 233.

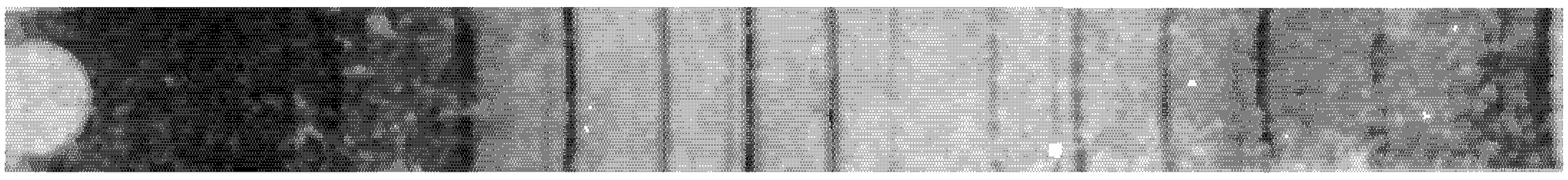


Рис. 233

Нелишне подчеркнуть, что, в отличие от лауэграмм, для получения которых требуется сплошное рентгеновское излучение, дебаеграммы получаются в *монохроматическом свете.* В белом свете никаких дебаеграмм с резкими дифракционными кольцами получиться не может.

10. Остановимся в заключение на вопросе о разрешающей способности дифракционной решетки (одномерной) в рентгеновской области спектра. Формула (47.3) для разрешающей способности решетки относится к случаю, когда решетка дает фраунгоферову дифракционную картину. Как было выяснено выше, в рентгеноскопии реализуется противоположный случай. Фокусирующего устройства нет, а фотопластинка ставится на малых расстояниях от решетки. В таких условиях дифракционные пучки подчиняются геометрической оптике, и формулой (47.3) пользоваться нельзя.

Пусть на решетку падает пучок параллельных лучей с длиной волны λ под углом скольжения α_0. Направление распространения дифрагированного пучка m-го порядка определится из условия $d(\cos\alpha_0 - \cos\alpha) = m\lambda$. Для такого же пучка с близкой длиной волны λ': $d(\cos\alpha_0 - \cos\alpha') = m\lambda'$. Отсюда $d(\cos\alpha' - \cos\alpha) = m(\lambda - \lambda')$, или

$$d\sin\alpha\cdot\delta\alpha = m\,\delta\lambda,$$

где введены обозначения: $\delta\alpha = |\alpha' - \alpha|$, $\delta\lambda = |\lambda' - \lambda|$. Для спектрального разрешения необходимо, чтобы оба пучка пространственно разделились. Если L — расстояние до фотопластинки, измеренное вдоль направления дифрагированного луча, то боковое смещение одного пучка относительно другого равно $x = L\delta\alpha$. Условие разрешения состоит в том, чтобы это смещение было не меньше ширины дифрагированного пучка: $x \geqslant h$. Последняя определяется выражением $h = D\sin\alpha$, где D — длина дифракционной решетки. В результате условие разрешения примет вид

$$\frac{Lm\,\delta\lambda}{d\sin\alpha} \geqslant D\sin\alpha.$$

Минимальному разрешаемому расстоянию соответствует знак равенства. Для него получаем

$$\delta\lambda = \frac{Dd\sin^2\alpha}{Lm}, \tag{61.10}$$

а для разрешающей способности

$$\frac{\lambda}{\delta\lambda} = \frac{Lm\lambda}{Dd\sin^2\alpha}. \tag{61.11}$$

Запишем эту формулу в виде

$$\frac{\lambda}{\delta\lambda} = Nm = \frac{L\lambda}{h^2}, \tag{61.12}$$

или

$$\frac{\lambda}{\delta\lambda} = \frac{Lm\lambda}{hd\sin\alpha}. \tag{61.13}$$

Для повышения разрешающей способности надо применять узкие пучки, а фотопластинку помещать как можно дальше. По сравнению с рэлеевской формулой формула (61.12) дает меньшее значение для разрешающей способности, поскольку она относится к области, в которой $L \ll h^2/\lambda$.

ГЛАВА V

ОТРАЖЕНИЕ И ПРЕЛОМЛЕНИЕ СВЕТА

§ 62. Поляризованный и естественный свет

1. При изучении интерференции и дифракции мы отвлеклись от поперечности световых колебаний, предполагая, что складываемые колебания совершаются в одном направлении. Обратимся теперь к изучению явлений *поляризации света*, типичных именно для *поперечных колебаний.* Плоская волна называется *линейно поляризованной* или *плоскополяризованной*, если электрический вектор **E** все время лежит в одной плоскости, в которой расположена также нормаль **N** к фронту волны (рис. 234). Эта плоскость называется *плоскостью колебаний* или *плоскостью поляризации*[1]. От поляризованного света следует отличать *естественный свет.* В нем в каждый момент времени векторы **E**, **H**, **N**, хотя и остаются взаимно перпендикулярными, но направления векторов **E** и **H** беспорядочно изменяются с течением времени. Поэтому естественный свет обладает (статистически) *осевой симметрией* относительно направления его распространения. Для линейно поляризованного света такой симметрии нет. Его свойства в различных плоскостях, проходящих через направление нормали **N**, различны. Среди этих плоскостей есть две избранные плоскости, в одной из которых лежит вектор **E**, а в другой — вектор **H**. Осевая асимметрия сохраняется и для смеси естественного света с линейно поляризованным. Такой свет называется *частично поляризованным.*

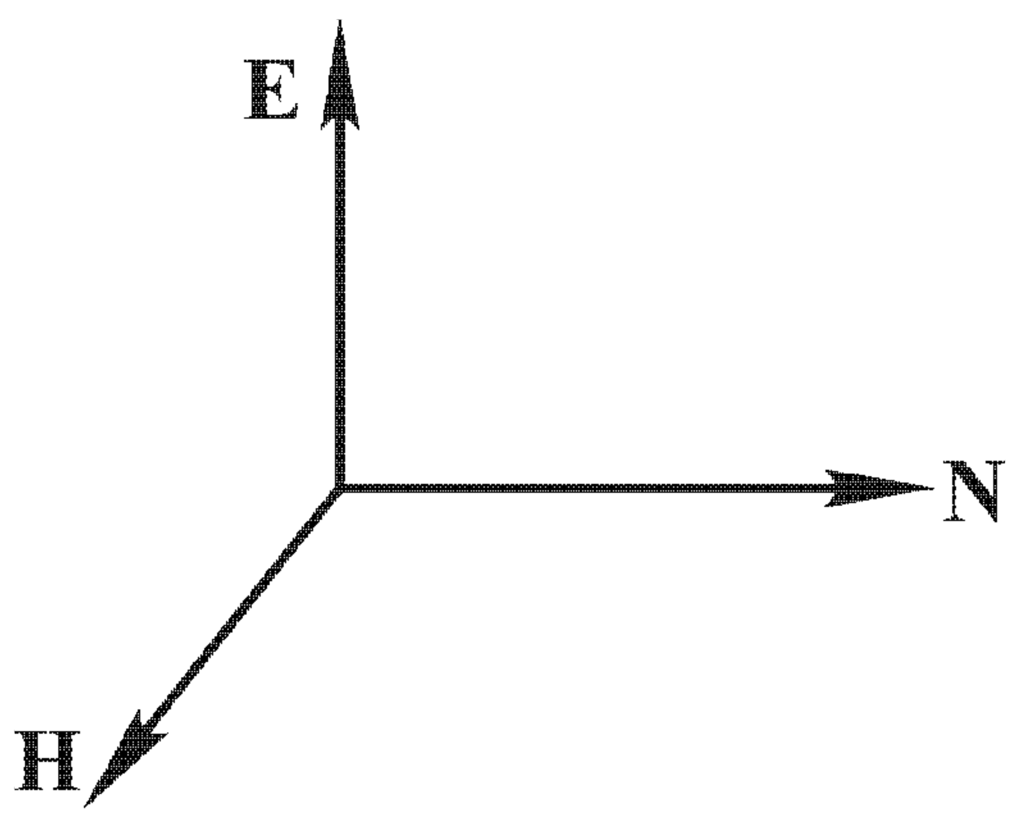


Рис. 234

---

[1] По старой терминологии плоскостью поляризации называлась плоскость $(\mathbf{N}, \mathbf{H})$, содержащая магнитный вектор **H**.

2. Линейно поляризованный свет легко получить, пропустив естественный свет через пластинку турмалина, вырезанную параллельно его кристаллографической (оптической) оси. Турмалин сильно поглощает световые лучи, в которых электрический вектор перпендикулярен к оптической оси. Если же электрический вектор параллелен оси, то такие лучи проходят через турмалин почти без поглощения. Поэтому естественный свет, пройдя через пластинку турмалина, наполовину поглощается и становится линейно поляризованным с электрическим вектором, ориентированным параллельно оптической оси турмалина.

Таким же свойством обладают *поляроиды*, более удобные в обращении. Они представляют собой искусственно приготовляемые коллоидные пленки, служащие для получения поляризованного света. Наиболее распространенным материалом для приготовления поляроидов является *герапатит*, представляющий собой соединение йода с хинином. Этот материал вводят в целлулоидную или желатиновую пленку. В ней ультрамикроскопические кристаллики герапатита каким-либо способом (обычно механически, например протаскиванием вязкой массы через узкую щель) ориентируются своими осями в одном и том же направлении. Полученная масса, подобно турмалину, действует как один кристалл и поглощает световые колебания, электрический вектор которых перпендикулярен к оптической оси. С другими способами получения поляризованного света мы познакомимся в дальнейшем.

Всякий прибор, служащий для получения поляризованного света, называется *поляризатором*. Тот же прибор, применяемый для исследования поляризации света, называется *анализатором*. Таким образом, кристаллы турмалина или поляроиды могут служить и поляризаторами, и анализаторами.

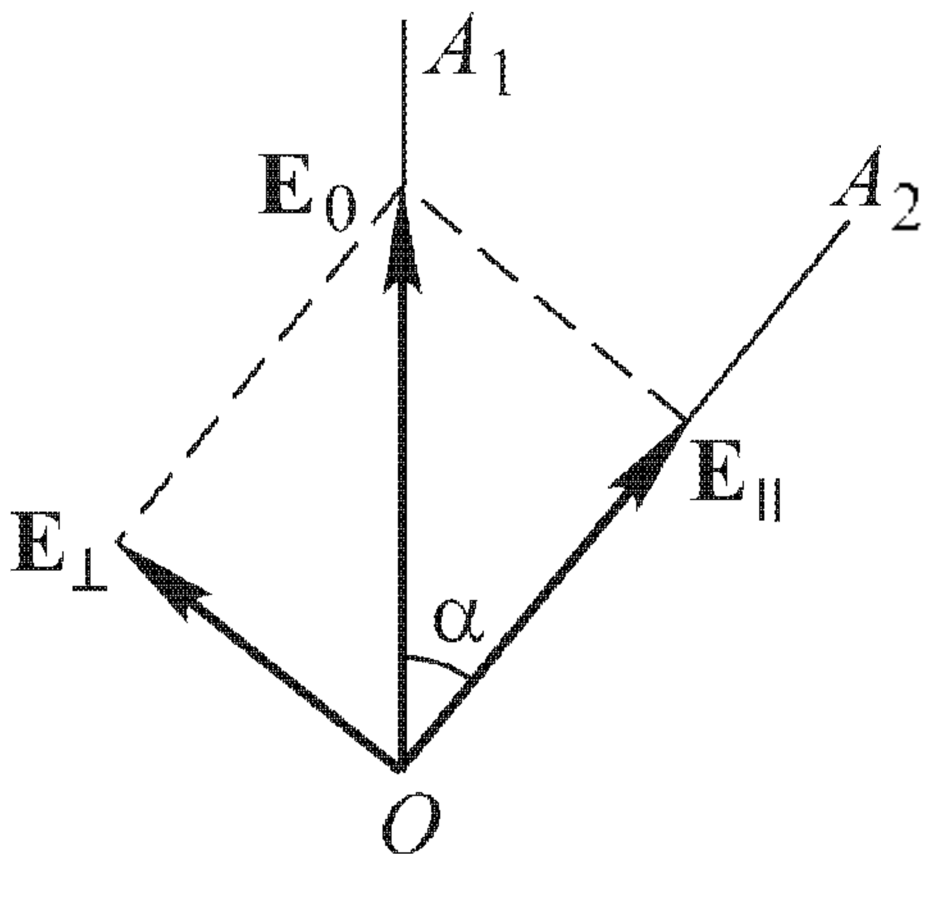


Рис. 235

3. Допустим, что два кристалла турмалина или два поляроида поставлены друг за другом, так что их оси OA_1 и OA_2 образуют между собой некоторый угол α (рис. 235). Первый поляроид пропустит свет, электрический вектор $\mathbf{E}_0$ которого параллелен его оси OA_1. Обозначим через I_0 интенсивность этого света. Разложим $\mathbf{E}_0$ на вектор $\mathbf{E}_{||}$, параллельный оси OA_2 второго поляроида, и вектор $\mathbf{E}_{\perp}$, перпендикулярный к ней ($\mathbf{E}_0 = \mathbf{E}_{||} + \mathbf{E}_{\perp}$). Составляющая $\mathbf{E}_{\perp}$ будет задержана вторым поляроидом. Через оба поляроида пройдет свет с электрическим вектором $\mathbf{E} \equiv \mathbf{E}_{||}$, длина которого $E = E_0 \cos\alpha$. Интенсивность света, прошедшего через оба поляроида, будет

$$I = I_0 \cos^2\alpha. \tag{62.1}$$

Такое соотношение справедливо для любого поляризатора и анализатора. Оно называется *законом Малюса* (1775–1812).

4. Важные состояния поляризации возникают при наложении монохроматических волн. Их общий характер одинаков для векторных волн любой физической природы. Для наглядности начнем с механического примера, когда частица совершает два гармонических колебания с одной и той же частотой ω: одно колебание происходит вдоль оси X, другое — вдоль оси Y. Координаты частицы представляются выражениями

$$x = a\cos\omega t, \quad y = b\cos(\omega t + \delta) \qquad (a, b > 0). \tag{62.2}$$

Исключив из этих выражений время t, найдем уравнение траектории частицы:

$$\frac{x^2}{a^2} - \frac{2xy}{ab}\cos\delta + \frac{y^2}{b^2} = \sin^2\delta. \tag{62.3}$$

Это — кривая второго порядка, а именно *эллипс*, так как координаты x и y, как это видно из (62.2), не могут обращаться в бесконечность. Таким образом, *от сложения двух взаимно перпендикулярных колебаний одной и той же частоты возникает движение по эллипсу.*

Для определения направления движения частицы по эллипсу заметим, что начальную фазу δ всегда можно выбрать так, чтобы она была заключена между $-\pi$ и $+\pi$. Тогда колебание с большей фазой называют *опережающим*, а с меньшей фазой — *запаздывающим.* Напишем далее компоненты скорости частицы вдоль осей X и Y:

$$\dot{x} = -\omega a\sin\omega t, \quad \dot{y} = -\omega b\sin(\omega t + \delta).$$

При $t = 0$ получим $x = a$, $\dot{y} = -\omega b\sin\delta$. Таким образом, при $t = 0$ частица находится на вертикальной прямой $x = a$. Если $-\pi < \delta < 0$, то $\dot{y} > 0$, т.е. частица движется вверх, описывая эллипс против часовой стрелки. Если же $0 < \delta < \pi$, то $\dot{y} < 0$, частица движется вниз, описывая эллипс по часовой стрелке (рис. 236). В обоих случаях *движение по эллипсу совершается от положительного конца оси опережающего колебания к положительному концу оси запаздывающего колебания.*

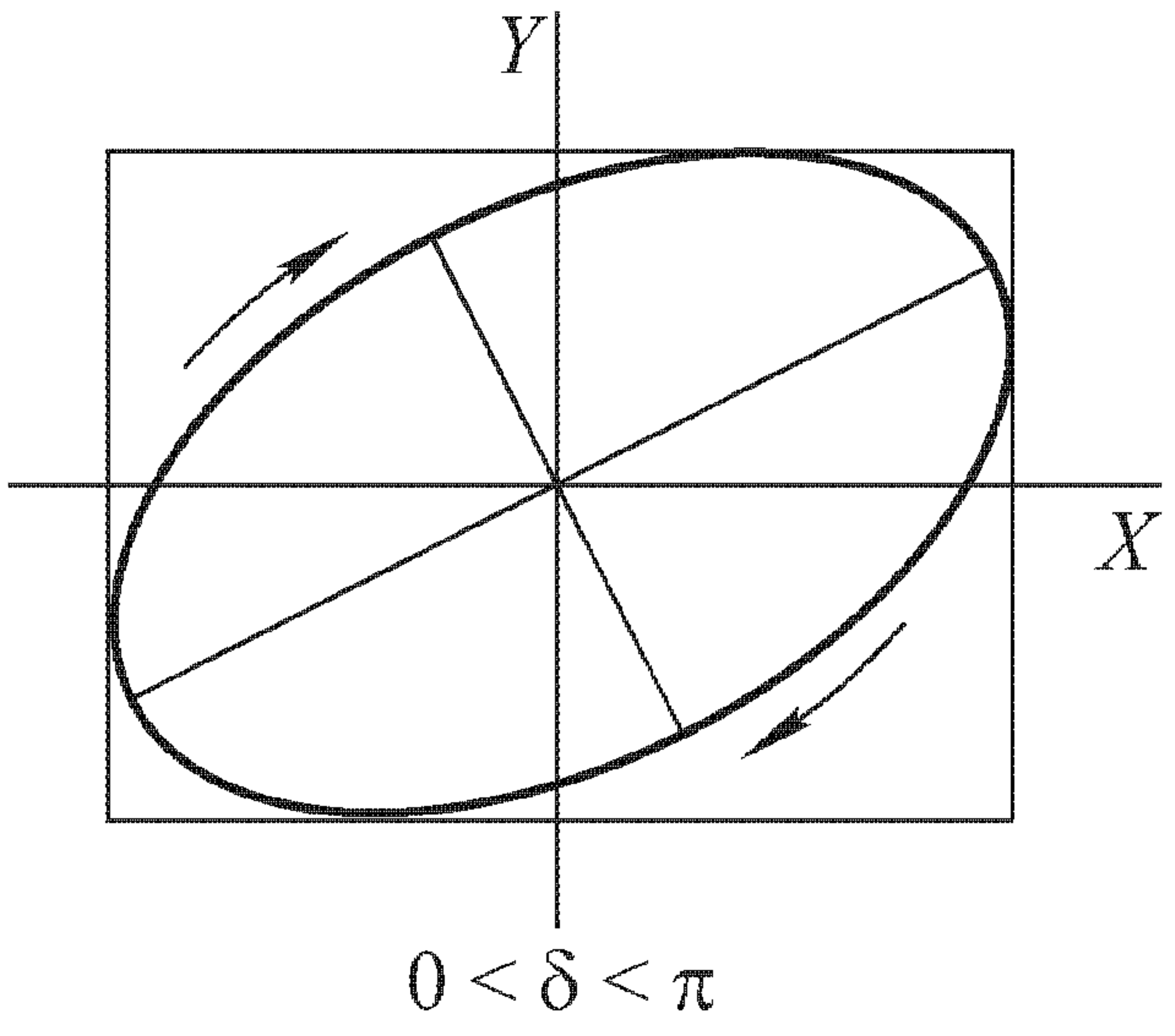


Рис. 236

В зависимости от значений параметров a, b, δ эллипс может вырождаться в отрезки прямой или в окружность. Из (62.3) видно, что для движения по окружности должны выполняться два условия: 1) $\cos\delta = 0$, т.е. $\delta = \pm\pi/2$; 2) $a = b$.

5. Все изложенное относится и к сложению *любых* векторных колебаний, в частности электромагнитных. Две электромагнитные волны, линейно поляризованные во взаимно перпендикулярных плоскостях, при сложении в общем случае дают волну, *поляризованную эллиптически.* В такой волне конец электрического (и магнитного) вектора в каждой точке пространства движется по эллипсу. Если эллипс вырождается в круг, то говорят, что волна *поляризована по кругу.*

Монохроматическое векторное поле всегда поляризовано, в общем случае эллиптически. Векторное поле называется монохроматическим, если все три его проекции на координатные оси совершают гармонические колебания с одной и той же частотой, т.е. представляются формулами вида

$$E_j = C_j(\mathbf{r})\cos[\omega t + \delta_j(\mathbf{r})] \qquad (j = x, y, z). \tag{62.4}$$

Умножая эти выражения на координатные орты $\mathbf{e}_j$ и суммируя по всем j, запишем монохроматическое поле в векторной форме:

$$\mathbf{E} = \mathbf{A}_1(\mathbf{r})\cos\omega t + \mathbf{A}_2(\mathbf{r})\sin\omega t. \tag{62.5}$$

Если векторы $\mathbf{A}_1(\mathbf{r})$ и $\mathbf{A}_2(\mathbf{r})$ везде или в некоторых точках коллинеарны, то в таких точках вектор $\mathbf{E}$ параллелен этим векторам, т.е. поле $\mathbf{E}$ поляризовано линейно. Если же $\mathbf{A}_1$ и $\mathbf{A}_2$ не коллинеарны, то, как видно из формулы (62.5), вектор $\mathbf{E}$ в любой момент времени лежит в плоскости векторов $\mathbf{A}_1(\mathbf{r})$ и $\mathbf{A}_2(\mathbf{r})$. Следовательно, конец вектора $\mathbf{E}$ описывает плоскую кривую. Чтобы найти ее форму, примем направление $\mathbf{A}_1$ за ось X, а перпендикулярное к нему направление, лежащее в плоскости $(\mathbf{A}_1, \mathbf{A}_2)$, — за ось Y. Тогда проекции E_x и E_y представятся в виде

$$E_x = a_x(\mathbf{r})\cos(\omega t + \delta_x), \quad E_y = a_y(\mathbf{r})\cos(\omega t + \delta_y).$$

Задача свелась к сложению двух взаимно перпендикулярных гармонических колебаний одной и той же частоты, сдвинутых по фазе относительно друг друга. От такого сложения получается движение по эллипсу.

Свет, испускаемый реальными источниками, *всегда не поляризован* или *поляризован частично.* Это является лучшим доказательством того, что он не монохроматичен.

§ 63. Число независимых граничных условий в электромагнитной теории света

1. *Формальная теория отражения и преломления света* строится на основе *граничных условий*, которым удовлетворяют векторы электромагнитного поля на границе раздела двух сред. Она определяет величины, характеризующие отраженную и преломленную волны, но ничего не говорит о *механизме* возникновения этих волн. На последний вопрос, а также на более тонкие вопросы дает ответ *молекулярная теория*. Сначала мы изложим формальную, а затем дадим краткое представление о молекулярной теории отражения и преломления света.

Будем рассматривать все тела как сплошные среды и предположим, что на границах раздела таких сред нет (в сущности, искусственно вводимых) поверхностных зарядов и токов. Тогда на границах раздела должны быть непрерывны тангенциальные составляющие векторов **E** и **H** и нормальные составляющие векторов **D** и **B**:

$$E_t^{(1)} = E_t^{(2)}, \quad H_t^{(1)} = H_t^{(2)}, \tag{63.1}$$

$$D_n^{(1)} = D_n^{(2)}, \quad B_n^{(1)} = B_n^{(2)}. \tag{63.2}$$

Все эти условия являются следствиями макроскопических уравнений Максвелла в интегральной форме, а потому они верны для *всяких сред*, пока последние можно рассматривать как сплошные. Условия (63.1) вытекают из уравнений

$$\begin{gathered} \oint \mathbf{E}\, d\mathbf{l} = -\frac{1}{c}\int \dot{\mathbf{B}}\, d\mathbf{F}, \\ \oint \mathbf{H}\, d\mathbf{l} = \frac{1}{c}\int \dot{\mathbf{D}}\, d\mathbf{F} + \frac{2\pi}{c}\mathcal{J}, \end{gathered} \tag{63.3}$$

а условия (63.2) — из уравнений

$$\oint \mathbf{D}\, d\mathbf{F} = 0, \qquad \oint \mathbf{B}\, d\mathbf{F} = 0 \tag{63.4}$$

(см. т. III, § 82).

2. Уравнения (63.1) и (63.2) не совсем независимы. Для исследования этого вопроса построим бесконечно короткий цилиндр, образующие которого перпендикулярны к границе раздела, а основания F_1 и F_2 лежат по разные стороны от нее (рис. 237). На основании теоремы о циркуляции вектора **H** получим

$$\oint\limits_{L_1} H_l^{(1)}\, dl = \frac{1}{c}\int \dot{D}_n^{(1)}\, dF_1,$$

$$\oint\limits_{L_2} H_l^{(2)}\, dl = \frac{1}{c}\int \dot{D}_n^{(2)}\, dF_2.$$

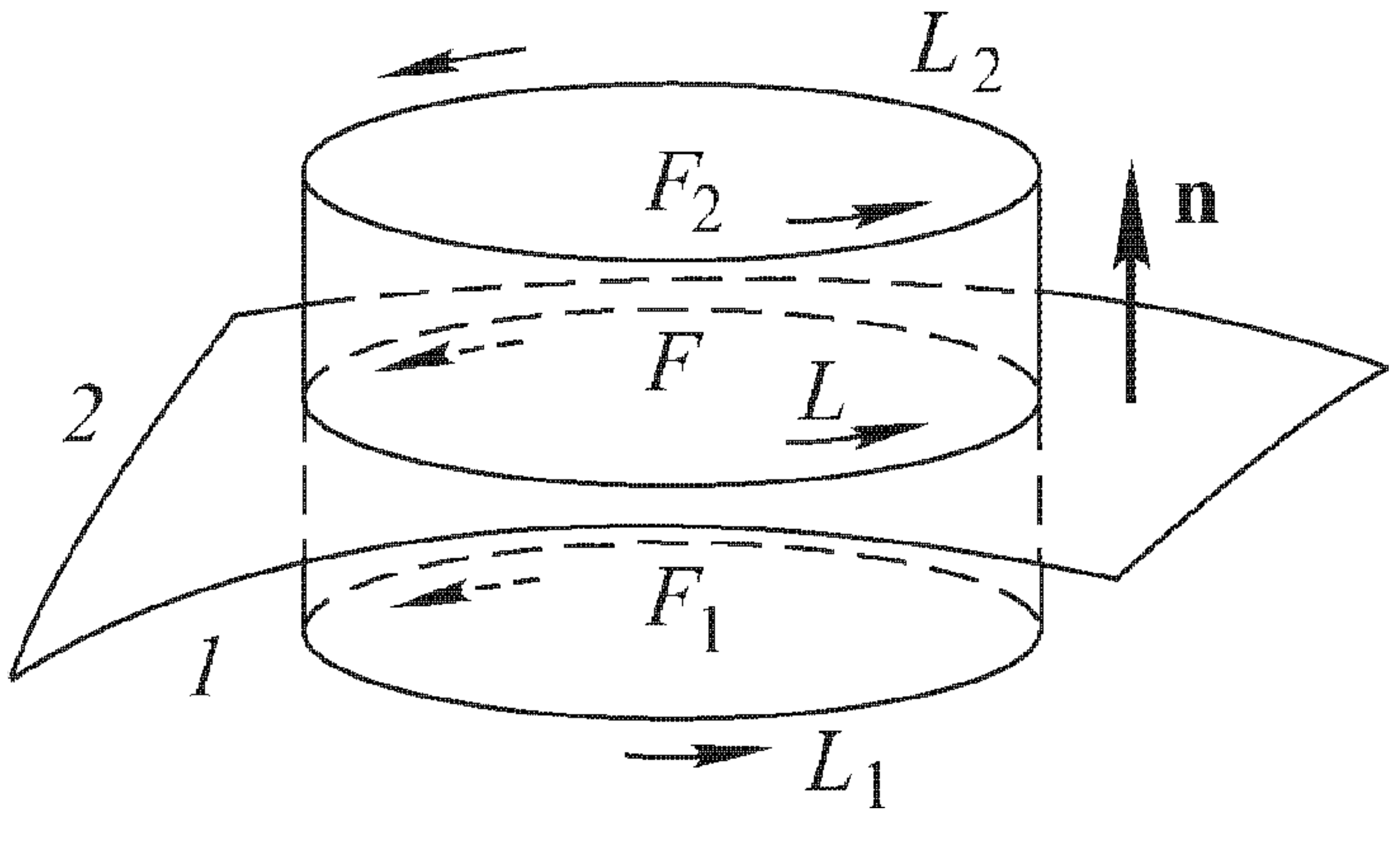


Рис. 237

В пределе, когда высота цилиндра обратится в нуль, контуры L_1 и L_2 сольются в общий контур L, а основания F_1 и F_2 — в общую площадку F, ограниченную контуром L. При этом, ввиду непрерывности тангенциальных составляющих вектора **H**, контурные интегралы совпадут между собой, а потому

$$\int\limits_F \dot{D}_n^{(1)}\,dF = \int\limits_F \dot{D}_n^{(2)}\,dF.$$

Аналогично,

$$\int\limits_F \dot{B}_n^{(1)}\,dF = \int\limits_F \dot{B}_n^{(2)}\,dF.$$

Отсюда, ввиду произвольности области интегрирования,

$$\dot{D}_n^{(1)} = \dot{D}_n^{(2)}, \qquad \dot{B}_n^{(1)} = \dot{B}_n^{(2)}. \tag{63.5}$$

Таким образом, граничные условия (63.5), как видно из их вывода, являются следствиями граничных условий (63.1) и уравнений Максвелла для циркуляций векторов **E** и **H**. В случае монохроматического поля $\dot{\mathbf{D}} = i\omega\mathbf{D}$, $\dot{\mathbf{B}} = i\omega\mathbf{B}$, так что условия (63.5) переходят в (63.2). Отсюда следует, что для монохроматических полей граничные условия (63.2) выполняются автоматически, если только выполняются условия (63.1). Поэтому в дальнейшем можно пользоваться только условиями (63.1), не заботясь о выполнении условий (63.2).

Если условия (63.1) записать в координатной форме, то получатся четыре уравнения, так как каждый из векторов **E** или **H** можно разложить на две тангенциальные и одну нормальную составляющие. Таким образом, электродинамика приводит к *четырем независимым граничным условиям.*

В старых теориях упругого эфира число независимых граничных условий было *шесть*: равенство трех составляющих смещений и трех составляющих сил упругих напряжений по обе стороны границы раздела. Чтобы удовлетворить этим шести граничным условиям, вообще говоря, необходимо, чтобы кроме поперечных волн существовали также и *продольные*. Но опыт говорил против существования продольных волн. Возникшую трудность теория пыталась устранить, наделяя эфир такими свойствами, чтобы продольные волны в нем никогда не возникали (несжимаемый или бесконечно сжимаемый эфир). Однако удовлетворительного решения проблемы таким путем получено не было. Электромагнитная теория не знает этой трудности, поскольку число независимых граничных условий в ней равно четырем. Им можно удовлетворить с помощью двух поперечных составляющих отраженной и двух поперечных составляющих преломленной волн.

§ 64. Геометрические законы отражения и преломления волн

1. Необходимость отражения и преломления света на границе раздела двух сред следует уже из граничных условий. Действительно, как будет видно из дальнейшего, граничные условия могут быть удовлетворены, вообще говоря, лишь при наличии отраженной и преломленной волн. Будем обозначать падающую волну индексом e (entfallende), отраженную — индексом r (reflektierte), проходящую — индексом d (durchgehende).

Пусть на плоскую неподвижную границу раздела падает плоская монохроматическая волна

$$\mathbf{E}^{(e)} = \mathcal{E} e^{i(\omega t - \mathbf{k}_1 \mathbf{r})}. \tag{64.1}$$

Из соображений симметрии следует, что отраженная и прошедшая волны

$$\mathbf{E}^{(r)} = \mathbf{R} e^{i(\omega t - \mathbf{k}'_1 \mathbf{r})}, \tag{64.2}$$

$$\mathbf{E}^{(d)} = \mathbf{D} e^{i(\omega t - \mathbf{k}_2 \mathbf{r})} \tag{64.3}$$

будут также плоскими и притом той же частоты ω. Равенство частот следует из линейности и однородности граничных условий. Если среды неподвижны, то коэффициенты при напряженностях полей в граничных условиях могут зависеть от координат, но не от времени. Пусть ω_r и ω_d — частоты отраженной и прошедшей волн. Тогда любое из граничных условий (63.1) принимает вид

$$A(\mathbf{r}) e^{i\omega t} + B(\mathbf{r}) e^{i\omega_r t} + C(\mathbf{r}) e^{i\omega_d t} = 0.$$

Коэффициенты $A(\mathbf{r})$, $B(\mathbf{r})$, $C(\mathbf{r})$ отличны от нуля, если только отраженная и прошедшая волны действительно существуют. Следовательно, функции $e^{i\omega t}$, $e^{i\omega_r t}$, $e^{i\omega_d t}$ линейно зависимы, а это возможно лишь при $\omega = \omega_r = \omega_d$. Если граница движется, то A, B и C зависят не только от $\mathbf{r}$, но и от времени. Тогда имеет место изменение частоты (эффект Доплера). В этой главе всюду предполагается, что среды неподвижны.

2. Найдем теперь волновые векторы отраженной и прошедшей волн. Формулы, определяющие эти векторы, называются *геометрическими законами отражения и преломления волн*. Они определяют *направления распространения отраженной и прошедшей волн*, а в случае их неоднородности также и *затухание в пространстве*.

Примем границу раздела сред за координатную плоскость XY. За ось X возьмем линию пересечения плоскости раздела сред с плоскостью падения. Ось Z направим вниз, т.е. в сторону второй среды. Тогда ось Y окажется перпендикулярной к плоскости падения и будет лежать в плоскости раздела сред. Так как по доказанному частоты падающей, отраженной и прошедшей волн одинаковы, то любое из граничных условий (63.1) примет вид

$$Ae^{-i(k_{1x}x+k_{1y}y)} + Be^{-i(k'_{1x}x+k'_{1y}y)} + Ce^{-i(k_{2x}x+k_{2y}y)} = 0,$$

где A, B, C — постоянные и притом отличные от нуля, если только отраженная и прошедшая волны действительно существуют. Полагая $y = 0$, получаем линейную зависимость между функциями $e^{-ik_{1x}x}$, $e^{-ik'_{1x}x}$, $e^{-ik_{2x}x}$ и поэтому заключаем, что

$$k_{1x} = k'_{1x} = k_{2x}. \tag{64.4}$$

Аналогично,

$$k_{1y} = k'_{1y} = k_{2y}. \tag{64.5}$$

Таким образом, тангенциальные составляющие волновых векторов отраженной и прошедшей волн равны тангенциальной составляющей волнового вектора падающей волны. Остается найти нормальные составляющие этих векторов. Согласно соотношению (5.14),

$$k_1'^2 = k_1^2 = \frac{\omega^2}{c^2}\varepsilon_1, \tag{64.6}$$

$$k_2^2 = \frac{\omega^2}{c^2}\varepsilon_2, \tag{64.7}$$

где ε_1 и ε_2 — диэлектрические проницаемости первой и второй сред. Далее

$$k'_{1z} = -\sqrt{k_1^2 - k_{1x}^2}, \tag{64.8}$$

$$k_{2z} = \sqrt{k_2^2 - k_{1x}^2}. \tag{64.9}$$

Знак минус перед корнем в формуле (64.8) взят потому, что плюс соответствует падающей волне. Что касается знака перед корнем в (64.9), то он будет определен в дальнейшем из физических соображений.

Если падающая волна однородна, то из (64.4), (64.5) и (64.8) следует, что отраженная волна также однородна. Ее волновая нормаль лежит в плоскости падения, а угол отражения равен углу падения. Для проходящей волны надо различать два случая.

Первый случай. $k_2^2 > k_{1x}^2$, т.е. преломленная волна однородна. Определим, какой знак следует выбрать в этом случае перед квадратным корнем в (64.9). Знаку плюс соответствует волна, распространяющаяся от границы раздела, — направление ее распространения обозначено на рис. 238 сплошной стрелкой. Знаку минус соответствует волна, идущая к границе раздела, — ее направление обозначено штриховой стрелкой.

Эти стрелки указывают направления распространения волновых фронтов, т.е. плоскостей равных фаз. Ясно, что отраженная и преломленная волны должны быть *уходящими* от границы раздела. Этим требованием обеспечивается однозначность решения задачи. Однако требование ухода относится не к фазе, а к энергии волны. Можно показать, что в случае электромагнитных волн в изотропных средах направления распространения фазы и энергии волны совпадают. Поэтому знак минус перед корнем в (64.9) следует отбросить; условиям задачи удовлетворяет только знак плюс.

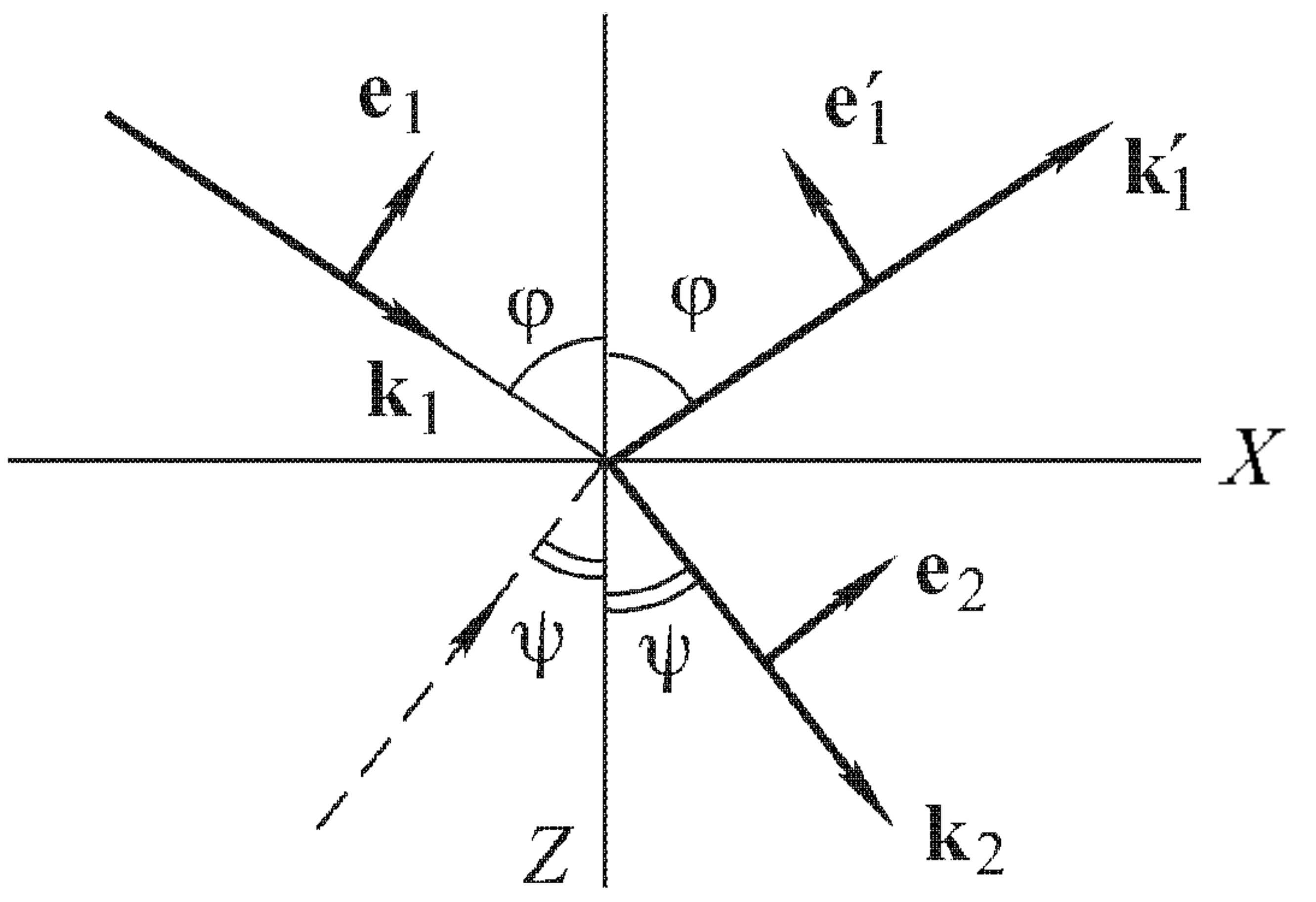


Рис. 238

Как видно из (64.4), нормали к падающей и преломленной волнам лежат в плоскости падения. Если φ — угол падения, а ψ — угол преломления, то

$$k_{1x} = k_1 \sin\varphi = \frac{\omega}{v_1}\sin\varphi, \qquad k_{2x} = k_2 \sin\psi = \frac{\omega}{v_2}\sin\psi,$$

откуда на основании (64.4)

$$\frac{\sin\varphi}{\sin\psi} = \frac{v_1}{v_2} = \frac{n_2}{n_1} = \sqrt{\frac{\varepsilon_2}{\varepsilon_1}}. \tag{64.10}$$

Второй случай. $k_2^2 < k_{1x}^2$, или $\omega^2/v_2^2 < \omega^2 \sin^2\varphi/v_1^2$, откуда $\sin\varphi > v_1/v_2 = n$. Здесь n — относительный показатель преломления второй среды относительно первой. Так как $\sin\varphi < 1$, то рассматриваемый случай возможен только при $n < 1$. Составляющая k_{2z} чисто мнимая, а волна во второй среде, если она существует, неоднородная. Знак корня в (64.9) определится из требования, чтобы при удалении от границы раздела амплитуда волны затухала. Этому требованию удовлетворяет только выражение

$$k_{2z} = -i\sqrt{k_{1x}^2 - k_2^2} = -\frac{i}{2h}. \tag{64.11}$$

В самом деле, тогда (64.3) принимает вид

$$\mathbf{E}^{(d)} = \mathbf{D}e^{-z/2h}e^{i(\omega t - k_{1x}x)}, \tag{64.12}$$

т.е. волна во второй среде будет затухать в направлении оси Z, чего не получилось бы, если бы в (64.11) вместо минуса взять плюс.

Плоскости равных фаз волны (64.12) перпендикулярны к оси X и распространяются вдоль нее с фазовой скоростью $v_x =$

$= \omega/k_{1x}$. Плоскости равных амплитуд параллельны границе раздела. При смещении вглубь среды на h интенсивность волны (пропорциональная квадрату амплитуды) убывает в e раз. Величина h называется *глубиной проникновения* волны во вторую среду. Она равна

$$h = \frac{1}{\sqrt{k_{1x}^2 - k_2^2}} = \frac{\lambda_1}{4\pi\sqrt{\sin^2\varphi - n^2}}, \tag{64.13}$$

где λ_1 — длина волны в первой среде.

Из (64.12) видно, что на больших (по сравнению с глубиной проникновения) расстояниях от границы раздела волна во второй среде практически полностью затухает. А так как поглощения света нет, то энергия падающей волны, проникшая во вторую среду, должна снова целиком возвратиться в первую среду. Иными словами, при $\sin\varphi > n$ отражение света должно быть *полным*. Угол φ_0, определяемый соотношением $\sin\varphi_0 = n$, называется *предельным углом полного отражения*.

3. В случае обыкновенного отражения

$$\begin{aligned} \sin\varphi &= k_{1x}/k_1, \quad \cos\varphi = k_{1z}/k_1, \\ \sin\psi &= k_{2x}/k_2, \quad \cos\psi = k_{2z}/k_2. \end{aligned} \tag{64.14}$$

В случае полного отражения не существует вещественного угла ψ, удовлетворяющего соотношениям (64.14), так как они дают для $\sin\psi$ значения, превосходящие единицу, а для $\cos\psi$ — мнимые значения. Однако в целях сохранения единой формы записи при обыкновенном и полном отражениях целесообразно сохранить формулы (64.14) как простые определения $\sin\psi$ и $\cos\psi$. Поскольку эти величины удовлетворяют соотношению $\sin^2\psi + \cos^2\psi = 1$, они могут рассматриваться как синус и косинус комплексного аргумента ψ в смысле теории функций комплексного переменного:

$$\sin\psi = \frac{e^{i\psi} - e^{-i\psi}}{2i}, \quad \cos\psi = \frac{e^{i\psi} + e^{-i\psi}}{2}.$$

Если $\sin\psi$ и $\cos\psi$ известны, то этими формулами аргумент ψ определяется с точностью до целого кратного от 2π. Это не может сказаться на однозначности физических выводов, так как во все формулы будет входить не сам комплексный угол ψ, а его синус и косинус. К так определенным функциям $\sin\psi$ и $\cos\psi$ применимы все формальные соотношения обычной тригонометрии. Поэтому над комплексными $\sin\psi$ и $\cos\psi$ можно выполнять все преобразования, как если бы они были обыкновенными синусом и косинусом.

Заметим, наконец, что вместо (64.14) можно написать

$$\sin\psi = \sin\varphi/n, \quad \cos\psi = -\frac{i}{n}\sqrt{\sin^2\varphi - n^2}. \tag{64.15}$$

§ 65. Формулы Френеля

1. При выводе геометрических законов отражения и преломления волн явный вид граничных условий не использовался. Для определения амплитуд отраженной и проходящей волн необходимо использовать граничные условия *в явном виде.*

Разложим электрическое поле каждой волны на две составляющие. Одна из них лежит в плоскости падения, другая перпендикулярна к этой плоскости. Часто эти составляющие называют *главными составляющими* соответствующих волн. Они обозначаются значками $\|$ и $\perp$ соответственно. Пусть $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$ — единичные векторы вдоль координатных осей, а $\mathbf{e}_1$, $\mathbf{e}_1'$, $\mathbf{e}_2$ — единичные векторы, лежащие в плоскости падения и перпендикулярные соответственно к падающему, отраженному и преломленному лучам (рис. 238). Тогда

$$\mathbf{e}_1 = \frac{[\mathbf{e}_y\mathbf{k}_1]}{k_1}, \quad \mathbf{e}_1' = \frac{[\mathbf{e}_y\mathbf{k}_1']}{k_1}, \quad \mathbf{e}_2 = \frac{[\mathbf{e}_y\mathbf{k}_2]}{k_2}. \tag{65.1}$$

В случае полного отражения вектор $\mathbf{e}_2$ — комплексный, его геометрическая интерпретация как единичного вектора, перпендикулярного к преломленному лучу, теряет смысл. Поэтому, чтобы охватить не только обыкновенное, но и полное отражение, дальнейшие вычисления проводятся чисто аналитически, не прибегая к чертежу. Введем разложения

$$\begin{aligned} \boldsymbol{\mathcal{E}} &= \mathcal{E}_\perp \mathbf{e}_y + \mathcal{E}_\| \mathbf{e}_1, \\ \mathbf{R} &= R_\perp \mathbf{e}_y + R_\| \mathbf{e}_1', \\ \mathbf{D} &= D_\perp \mathbf{e}_y + D_\| \mathbf{e}_2 \end{aligned} \tag{65.2}$$

и определим амплитуды $R_\perp$, $R_\|$, $D_\perp$, $D_\|$. Умножая скалярно первое из уравнений (65.2) на $\mathbf{e}_x$, находим

$$\mathcal{E}_x = \mathcal{E}_\|(\mathbf{e}_1\mathbf{e}_x) = \frac{\mathcal{E}_\|}{k_1}(\mathbf{e}_x[\mathbf{e}_y\mathbf{k}_1]) = \frac{\mathcal{E}_\|}{k_1}([\mathbf{e}_x\mathbf{e}_y]\mathbf{k}_1) = \frac{\mathcal{E}_\|}{k_1}(\mathbf{e}_z\mathbf{k}_1) = \mathcal{E}_\| \cos\varphi.$$

Аналогично, $\mathcal{E}_y = \mathcal{E}_\perp$, $\mathcal{E}_z = -\mathcal{E}_\| \sin\varphi$. Таким образом, x-, y-, z-составляющие электрического поля на границе раздела сред (т.е. при $z = 0$) будут

$$E_x^{(e)} = \cos\varphi \cdot \mathcal{E}_\|, \quad E_y^{(e)} = \mathcal{E}_\perp, \quad E_z^{(e)} = -\sin\varphi \cdot \mathcal{E}_\|. \tag{65.3}$$

При этом мы всюду опустили фазовые множители, так как в любой точке границы раздела они одинаковы для всех трех волн: падающей, отраженной и прошедшей. Магнитные свойства сред учитывать не будем, считая, что $\mathbf{H} \equiv \mathbf{B}$. Магнитное поле вычисляем по формуле (5.6) и находим

$$H_x^{(e)} = -n_1 \cos\varphi \cdot \mathcal{E}_\perp, \quad H_y^{(e)} = n_1 \mathcal{E}_\|, \quad H_z^{(e)} = n_1 \sin\varphi \cdot \mathcal{E}_\perp. \tag{65.3a}$$

Для отраженной волны:

$$\begin{gathered} E_x^{(r)} = -\cos\varphi\cdot R_{\parallel}, \qquad E_y^{(r)} = R_\perp, \qquad E_z^{(r)} = -\sin\varphi\cdot R_{\parallel}, \\ H_x^{(r)} = n_1\cos\varphi\cdot R_\perp, \quad H_y^{(r)} = n_1 R_{\parallel}, \quad H_z^{(r)} = n_1\sin\varphi\cdot R_\perp. \end{gathered} \tag{65.4}$$

Для прошедшей волны:

$$\begin{gathered} E_x^{(d)} = \cos\psi\cdot D_{\parallel}, \qquad E_y^{(d)} = D_\perp, \qquad E_z^{(d)} = -\sin\psi\cdot D_{\parallel}, \\ H_x^{(d)} = -n_2\cos\psi\cdot D_\perp, \quad H_y^{(d)} = n_2 D_{\parallel}, \quad H_z^{(d)} = D_{\parallel} n_2\sin\psi. \end{gathered} \tag{65.5}$$

Для определения четырех неизвестных $R_\perp$, $R_\parallel$, $D_\perp$, $D_\parallel$ электродинамика дает четыре независимых граничных условия:

$$\begin{gathered} E_x^{(e)} + E_x^{(r)} = E_x^{(d)}, \quad E_y^{(e)} + E_y^{(r)} = E_y^{(d)}, \\ H_x^{(e)} + H_x^{(r)} = H_x^{(d)}, \quad H_y^{(e)} + H_y^{(r)} = H_y^{(d)}. \end{gathered} \tag{65.6}$$

Подставляя в них найденные выше значения, получим

$$\begin{gathered} \cos\varphi\cdot(\mathcal{E}_\parallel - R_\parallel) = \cos\psi\cdot D_\parallel, \qquad \mathcal{E}_\perp + R_\perp = D_\perp, \\ n_1\cos\varphi\cdot(\mathcal{E}_\perp - R_\perp) = n_2\cos\psi\cdot D_\perp, \quad n_1(\mathcal{E}_\parallel + R_\parallel) = n_2 D_\parallel. \end{gathered}$$

Отсюда

$$\begin{gathered} r_\perp \equiv \frac{R_\perp}{\mathcal{E}_\perp} = \frac{n_1\cos\varphi - n_2\cos\psi}{n_1\cos\varphi + n_2\cos\psi}, \quad d_\perp \equiv \frac{D_\perp}{\mathcal{E}_\perp} = \frac{2n_1\cos\varphi}{n_1\cos\varphi + n_2\cos\psi}, \\ r_\parallel \equiv \frac{R_\parallel}{\mathcal{E}_\parallel} = \frac{n_2\cos\varphi - n_1\cos\psi}{n_2\cos\varphi + n_1\cos\psi}, \quad d_\parallel \equiv \frac{D_\parallel}{\mathcal{E}_\parallel} = \frac{2n_1\cos\varphi}{n_2\cos\varphi + n_1\cos\psi}. \end{gathered} \tag{65.7}$$

Отношения $R_\perp/\mathcal{E}_\perp$, $D_\perp/\mathcal{E}_\perp$, ... называются *коэффициентами Френеля*. С помощью соотношений (64.10) формулы (65.7) легко привести к виду

$$\begin{gathered} \frac{R_\perp}{\mathcal{E}_\perp} = -\frac{\sin(\varphi-\psi)}{\sin(\varphi+\psi)}, \qquad \frac{D_\perp}{\mathcal{E}_\perp} = \frac{2\cos\varphi\sin\varphi}{\sin(\varphi+\psi)}, \\ \frac{R_\parallel}{\mathcal{E}_\parallel} = \frac{\operatorname{tg}(\varphi-\psi)}{\operatorname{tg}(\varphi+\psi)}, \quad \frac{D_\parallel}{\mathcal{E}_\parallel} = \frac{2\cos\varphi\sin\varphi}{\sin(\varphi+\psi)\cos(\varphi-\psi)}. \end{gathered} \tag{65.8}$$

Эти формулы и решают поставленную задачу. Они были впервые получены в 1821 г. Френелем из механической теории упругого эфира с помощью весьма простых, но нестрогих и противоречивых рассуждений. Последующие попытки строгого решения задачи с помощью уравнений упругости ухудшили дело: в случае волн, поляризованных в плоскости падения, они приводили к формулам, не согласующимся с опытом. Только электромагнитная теория света впервые дала строгое и удовлетворительное решение задачи.

2. При обыкновенном отражении угол ψ, а с ним и все коэффициенты Френеля *вещественны*. Отсюда следует, что отражение и преломление *не сопровождаются изменением фаз*, за исключением, быть может, изменения фазы отраженной волны на 180°. *Если падающая волна была линейно поляризована, то отраженная и преломленная волны будут также линейно поляризованы.*

При нормальном падении волны формулы (65.7) дают

$$\frac{R_\perp}{\mathcal{E}_\perp} = -\frac{R_\parallel}{\mathcal{E}_\parallel} = \frac{n_1 - n_2}{n_1 + n_2}. \tag{65.9}$$

В этом случае исчезает разница между волнами, поляризованными в плоскости падения и перпендикулярно к ней. Различие в знаках коэффициентов Френеля $R_\perp/\mathcal{E}_\perp$ и $R_\parallel/\mathcal{E}_\parallel$ связано с условным выбором положительных направлений, в которых проведены векторы $\mathbf{e}_1$ и $\mathbf{e}_1'$: в предельном случае нормального падения эти векторы направлены прямо противоположно (см. рис. 238). Если $n_1 < n_2$, то $R_\perp/\mathcal{E}_\perp < 0$; если $n_1 > n_2$, то $R_\perp/\mathcal{E}_\perp > 0$. *В первом случае отражение света сопровождается скачкообразным изменением фазы на* 180°. *Во втором оно происходит без изменения фазы.*

При скользящем падении, когда угол φ близок к 90°,

$$\frac{R_\perp}{\mathcal{E}_\perp} = \frac{R_\parallel}{\mathcal{E}_\parallel} \to -1,$$

т.е. отражение практически полное. С этим связано происхождение ярких и красивых изображений противоположных берегов в спокойных водоемах. Этим же объясняется, почему изображение заходящего солнца в тех же условиях по яркости почти не уступает самому солнцу.

Как видно из формул (65.7) или (65.8), отношение $R_\perp/\mathcal{E}_\perp$ никогда не обращается в нуль, за исключением тривиального случая $n_1 = n_2$. Напротив, при $\varphi + \psi = \pi/2$ знаменатель $\operatorname{tg}(\varphi + \psi)$ обращается в бесконечность. В этом случае $R_\parallel = 0$, т.е. отражение отсутствует. Таким образом, если электрический вектор падающей волны лежит в плоскости падения, то при некотором угле падения отражение света исчезает. В этом состоит *закон Брюстера* (1781–1868), открытый им экспериментально в 1815 г. Соответствующий угол падения φ_B называется *углом Брюстера*. Для его вычисления заметим, что при $\varphi + \psi = \pi/2$ направления прошедшего и отраженного лучей взаимно перпендикулярны (рис. 239). Следовательно, $\cos\varphi_B = \sin\psi_B = \frac{1}{n}\sin\varphi_B$, т.е.

$$\operatorname{tg}\varphi_B = n. \tag{65.10}$$

Если неполяризованный свет падает под углом Брюстера, то *составляющая с электрическим вектором* $\mathbf{E}_{\parallel}$ *отражаться не будет. Отраженный свет окажется линейно поляризованным и притом перпендикулярно к плоскости падения*[1].

Отражение под углом Брюстера дает простейший способ получения поляризованного света. Недостаток этого способа — малая интенсивность отраженного света. Для его устранения применяют несколько стеклянных пластинок, сложенных в стопу (стопу Столетова). При большом числе пластинок отражается почти половина падающего света. Если свет падает под углом Брюстера, то отраженный свет оказывается полностью поляризованным перпендикулярно к плоскости падения. Прошедший свет поляризован в плоскости падения, но эта поляризация не совсем полная. В настоящее время стопа для получения поляризованного света почти не употребляется, так как существуют более удобные и совершенные способы (*поляризационные призмы, поляроиды*).

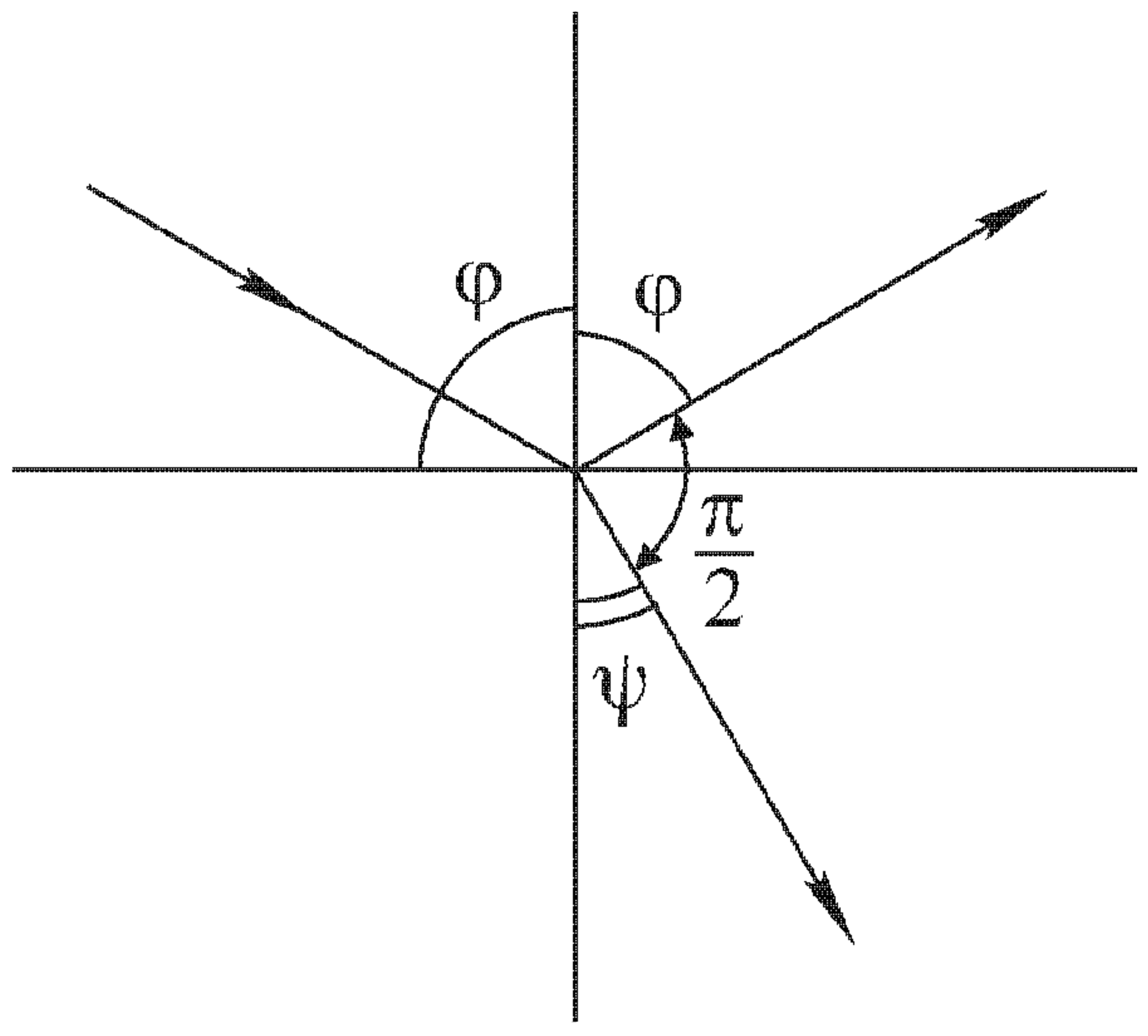


Рис. 239

3. Отношение отраженной энергии к энергии падающей называется *коэффициентом отражения*. Так как энергия пропорциональна квадрату амплитуды, то формулы (65.7) дают для коэффициентов отражения главных составляющих падающей волны следующие выражения:

$$\rho_{\perp} = \left(\frac{\cos\varphi - n\cos\psi}{\cos\varphi + n\cos\psi}\right)^2, \quad \rho_{\parallel} = \left(\frac{n\cos\varphi - \cos\psi}{n\cos\varphi + \cos\psi}\right)^2. \tag{65.11}$$

Коэффициент отражения при нормальном падении R называется *отражательной способностью*:

$$R = \left(\frac{n-1}{n+1}\right)^2. \tag{65.12}$$

Для стекла ($n = 1{,}5$) $R = 0{,}04 = 4\,\%$; для воды ($n = 1{,}33$) $R = 2\,\%$. При отражении от воды длинных электромагнитных волн ($n = \sqrt{\varepsilon} = 9$) $R = 64\,\%$. Эти цифры показывают, что ни вода, ни стекло при нормальном падении не могут служить зеркалом. Обычные зеркала используют отражение света от металлических поверхностей. Стекло служит только для защиты их задних посеребренных поверхностей. Однако наличие даже слабого отражения от передней стороны стекла делает такие посеребренные с задней стороны зеркала непригодными для оптических целей. Для этих целей необходимо покрывать металлом (лучше всего родием) переднюю поверхность стекла.

---

[1] Малюс, открывший поляризацию света при отражении, назвал плоскость падения плоскостью поляризации отраженного света. Здесь под *плоскостью поляризации* понимается плоскость, перпендикулярная к плоскости поляризации Малюса.

Отношение b прошедшей энергии к энергии падающей называется *коэффициентом пропускания*. Коэффициент пропускания можно также определить как отношение нормальных компонент усредненного по времени вектора Пойнтинга прошедшей и падающей волн. На основании (65.3) и (65.3а) находим для падающей волны

$$S_z^{(e)} = \frac{c}{16\pi}[E_x^{(e)}H_y^{*(e)} - E_y^{(e)}H_x^{*(e)}] + \text{компл. сопр.} = \\ = \frac{cn_1}{8\pi}\cos\varphi(|\mathcal{E}_\perp|^2 + |\mathcal{E}_\parallel|^2),$$

а для прошедшей

$$S_z^{(d)} = \frac{cn_2}{8\pi}\cos\psi(|D_\perp|^2 + |D_\parallel|^2).$$

Предполагая сначала, что вектор $\mathbf{E}$ перпендикулярен к плоскости падения, а затем параллелен ей, находим

$$b_\perp = \frac{n\cos\psi}{\cos\varphi}\left(\frac{2\cos\varphi}{\cos\varphi + n\cos\psi}\right)^2, \quad b_\parallel = \frac{n\cos\psi}{\cos\varphi}\left(\frac{2\cos\varphi}{n\cos\varphi + \cos\psi}\right)^2. \tag{65.13}$$

Нетрудно проверить, что $\rho_\perp + b_\perp = \rho_\parallel + b_\parallel = 1$, как и должно быть согласно закону сохранения энергии.

Коэффициент пропускания при нормальном падении

$$B = \frac{4n}{(n+1)^2} \tag{65.14}$$

называется *поверхностной прозрачностью*. Измерения отражательной способности и поверхностной прозрачности дают удобный метод измерения показателей преломления в инфракрасной части спектра.

Найдем теперь коэффициент отражения для случая, когда плоскость поляризации падающей волны составляет с плоскостью падения угол α. Этот угол называется *азимутом поляризации* падающей волны. Очевидно, $\mathcal{E}_\parallel = A\cos\alpha$, $\mathcal{E}_\perp = A\sin\alpha$, где A — амплитуда падающей волны. Полная отраженная энергия пропорциональна $\rho_\parallel\mathcal{E}_\parallel^2 + \rho_\perp\mathcal{E}_\perp^2 = A^2(\rho_\parallel\cos^2\alpha + \rho_\perp\sin^2\alpha)$. Следовательно, коэффициент отражения равен

$$\rho_\alpha = \rho_\parallel\cos^2\alpha + \rho_\perp\sin^2\alpha. \tag{65.15}$$

Коэффициент отражения неполяризованного света ρ получается отсюда путем усреднения по α. Так как в неполяризованном свете все направления электрического (и магнитного) вектора представлены с одинаковой вероятностью, то $\overline{\cos^2\alpha} = \overline{\sin^2\alpha} = 1/2$. Поэтому

$$\rho = \frac{1}{2}(\rho_{\|} + \rho_{\perp}). \tag{65.16}$$

На рис. 240 приведены теоретические кривые для коэффициентов отражения видимого света от стекла ($n = 1{,}5$), а на

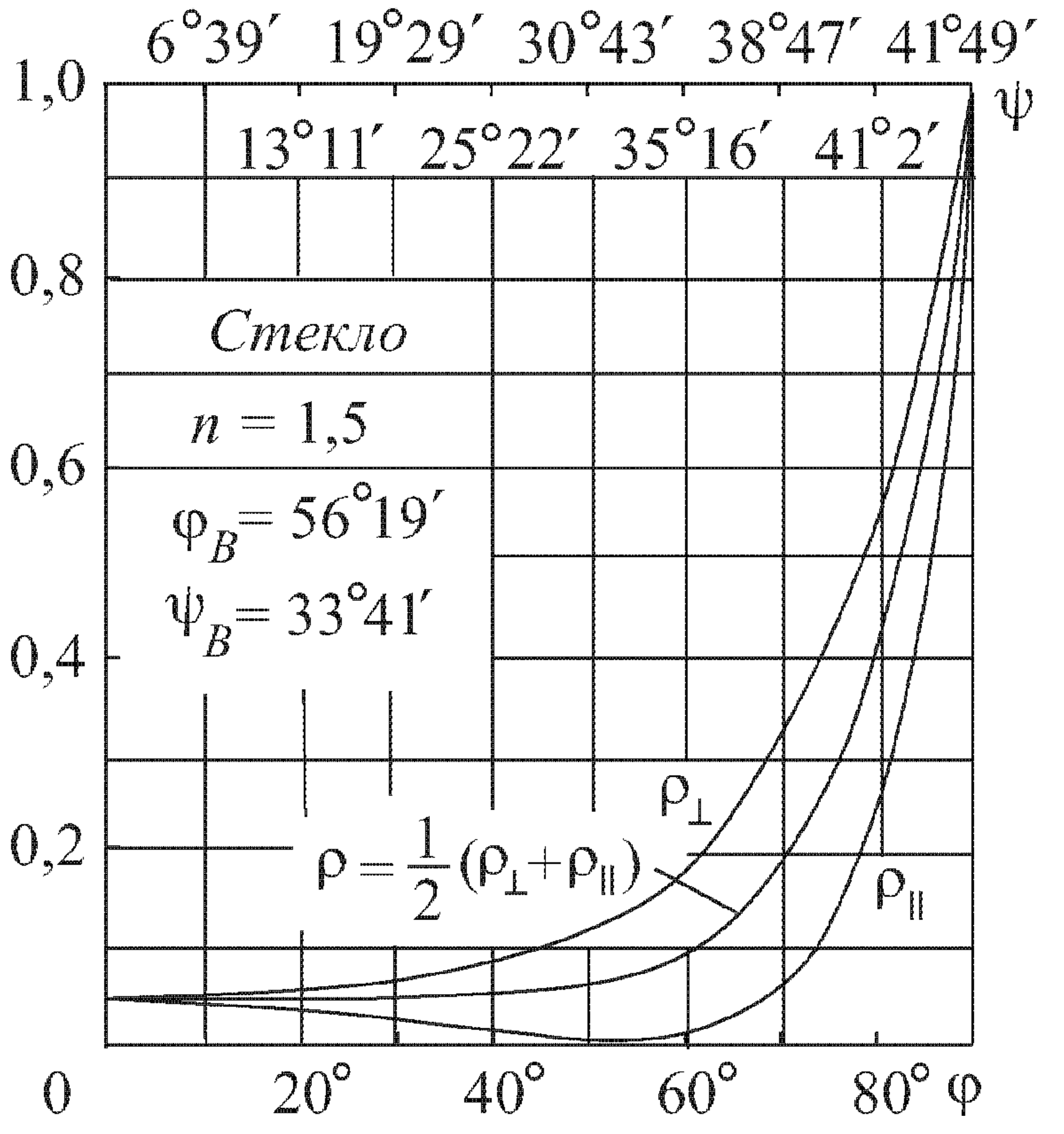


Рис. 240

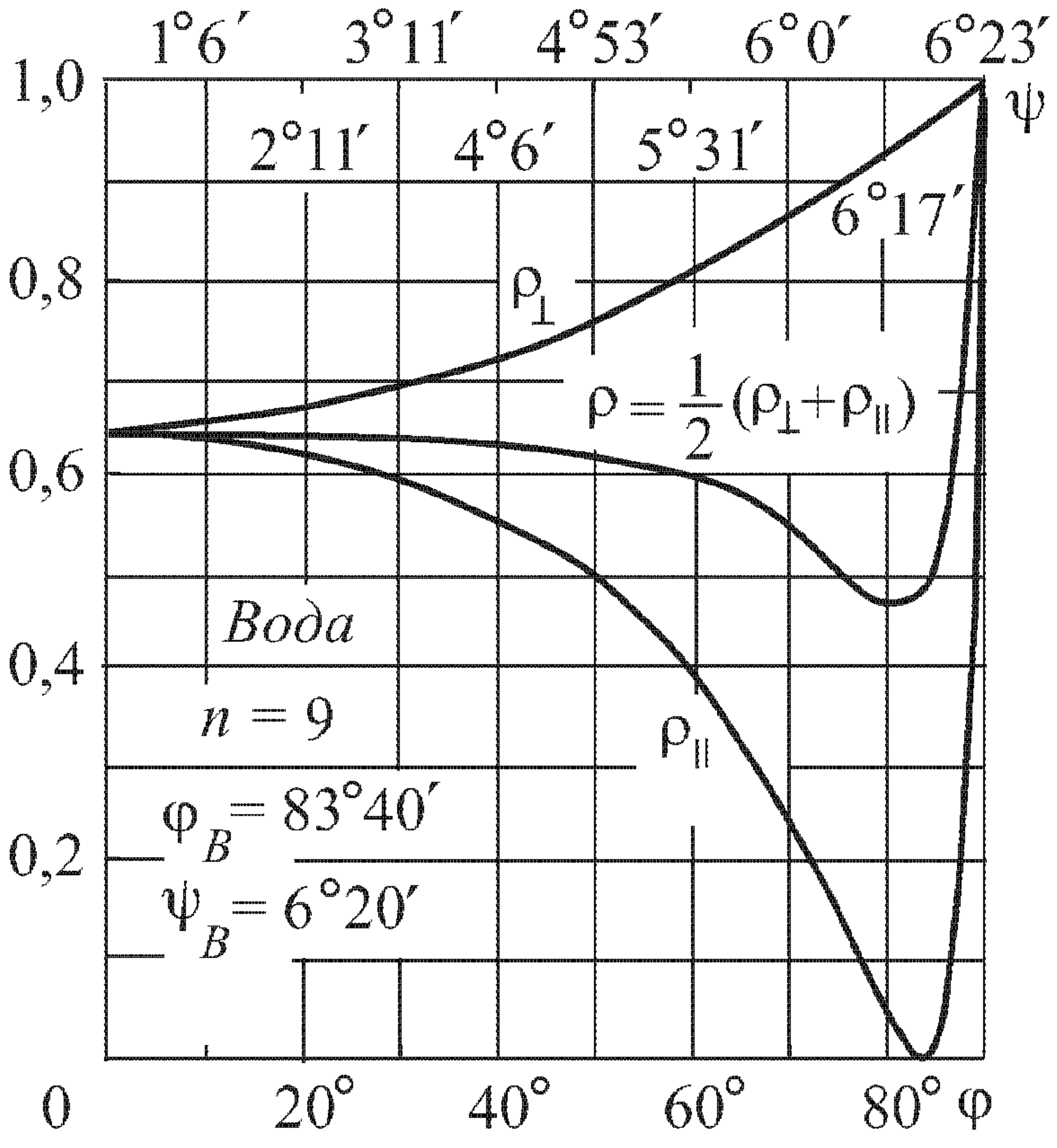


Рис. 241

рис. 241 — для радиоволн при их отражении от поверхности воды ($n = \sqrt{\varepsilon_0} = 9$)

Многочисленные измерения коэффициентов отражения при различных углах падения и при различных поляризациях падающей волны, предпринимавшиеся с целью проверки формул Френеля, находятся в очень хорошем согласии с этими формулами как в случае видимого света, так и в случае радиоволн. Исключение составляет случай отражения под углом Брюстера и в его ближайшей окрестности, где наблюдаются незначительные отступления от формул Френеля (см. § 70).

4. Формулы Френеля мы вывели в предположении, что свет *монохроматический*. Однако в случае обыкновенного отражения в эти формулы не входит длина волны, а отражение не сопровождается изменением фазы. Поэтому в случае недиспергирующих сред и обыкновенного отражения формулы Френеля справедливы и для *немонохроматических волн*. Надо только под $\mathcal{E}_{\perp}$, $R_{\perp}$, $D_{\perp}$ и т.д. понимать соответствующие компоненты напряженностей электрического поля падающей, отраженной и прошедшей волн *на границе раздела*. Это непосредственно следует из теоремы Фурье и принципа суперпозиции.

§ 66. Полное отражение

1. В случае полного отражения $\cos\psi$ — чисто мнимая величина, при этом формулы Френеля удобно записать в виде

$$\frac{R_\perp}{\mathcal{E}_\perp} = \frac{\cos\varphi + i\sqrt{\sin^2\varphi - n^2}}{\cos\varphi - i\sqrt{\sin^2\varphi - n^2}}, \quad \frac{D_\perp}{\mathcal{E}_\perp} = \frac{2\cos\varphi}{\cos\varphi - i\sqrt{\sin^2\varphi - n^2}},$$
$$\frac{R_\parallel}{\mathcal{E}_\parallel} = \frac{n^2\cos\varphi + i\sqrt{\sin^2\varphi - n^2}}{n^2\cos\varphi - i\sqrt{\sin^2\varphi - n^2}}, \quad \frac{D_\parallel}{\mathcal{E}_\parallel} = \frac{2n\cos\varphi}{n^2\cos\varphi - i\sqrt{\sin^2\varphi - n^2}}. \tag{66.1}$$

Из этих формул видно, что $|R_\perp| = |\mathcal{E}_\perp|$ и $|R_\parallel| = |\mathcal{E}_\parallel|$, т.е. отражение действительно является «полным». В то же время неоднородная волна во второй среде, вообще говоря, не исчезает. Например, если $\varphi = \varphi_0 = \arcsin n$, то $D_\perp = 2\mathcal{E}_\perp$. Здесь нет нарушения закона сохранения энергии. Дело в том, что формулы Френеля относятся к *монохроматическому полю*, т.е. к некоторому *установившемуся процессу*. А в этом случае закон сохранения энергии требует только, чтобы количество энергии во второй среде менялось во времени *периодически*. Иными словами, *среднее за период изменение энергии во второй среде должно быть равно нулю*. Что это условие действительно выполняется, показывает следующее простое вычисление.

Вычислим усредненную по времени нормальную составляющую вектора Пойнтинга во второй среде. Пусть электрический вектор лежит в плоскости падения. Тогда

$$\overline{S}_z = \frac{c}{16\pi} E_x^* H_y + \text{компл. сопр.}$$

Поле во второй среде определяется формулами (65.5). Так как $\cos\psi$ — величина чисто мнимая, то из (65.5) видно, что между электрическим и магнитным полями существует сдвиг фаз в 90°. При этом величина $E_x^* H_y$ также чисто мнимая, а потому $\bar{S}_z = 0$. То же справедливо и в случае, когда электрический вектор перпендикулярен к плоскости падения. Этим наше утверждение доказано.

Вычисление остальных компонент усредненного вектора Пойнтинга дает

$$\bar{S}_x = \frac{cn_1}{8\pi}(|D_\perp|^2 + |D_\parallel|^2)\sin\varphi e^{-z/h}, \quad \bar{S}_y = 0, \tag{66.2}$$

где h — глубина проникновения, определяемая выражением (64.13). Таким образом, во второй среде действительно имеется энергия, средний запас которой со временем не изменяется, а средний поток параллелен оси X. Существование такой энергии при установившемся режиме, очевидно, не противоречит полному отражению падающей волны.

Чтобы ответить на вопрос, как появилась энергия во второй среде, надо было бы исследовать *процесс установления колебаний.* Можно, например, рассмотреть квазимонохроматическую волну с передовым фронтом, перед которым нет никакого волнового возмущения. Пока фронт волны не достиг границы раздела, во второй среде нет поля. Как только волна дойдет до границы раздела, она сначала будет почти целиком проникать во вторую среду и лишь частично отражаться. По мере установления колебаний коэффициент отражения будет быстро нарастать и стремиться к своему предельному значению — единице. Полное отражение имеет место лишь для установившегося режима. Пока процесс не установился, отражение *всегда частичное.*

2. Пусть две одинаковые среды разделены тонким воздушным промежутком, на который падает волна под углом, бóльшим предельного. Можно ожидать (см. § 67), что она проникнет в воздушный зазор в виде неоднородной волны. При достаточно малой толщине зазора неоднородная волна достигнет его второй границы еще не очень сильно ослабленной. Вступив из воздушного зазора во вторую среду, волна снова превратится в однородную и может быть обнаружена обычными средствами. Опыт такого рода был выполнен еще Ньютоном и рассматривался как экспериментальное доказательство проникновения света в оптически менее плотную среду при полном отражении. (Строго говоря, здесь нет полного отражения: оно возмущено наличием второй среды.) Ньютон прижимал к гипотенузной грани прямоугольной призмы другую призму, сошлифованную сферически (рис. 242). Оказалось, что свет проходит во вторую призму не только в месте соприкосновения (оптический контакт), но и в небольшом кольце вокруг него — там, где толщина воздушного промежутка сравнима с длиной волны. При наблюдении в белом свете внешний край кольца имел красноватую окраску. Этого и следовало ожидать, так как глубина проникновения пропорциональна длине волны: для красных лучей она больше, чем для синих.

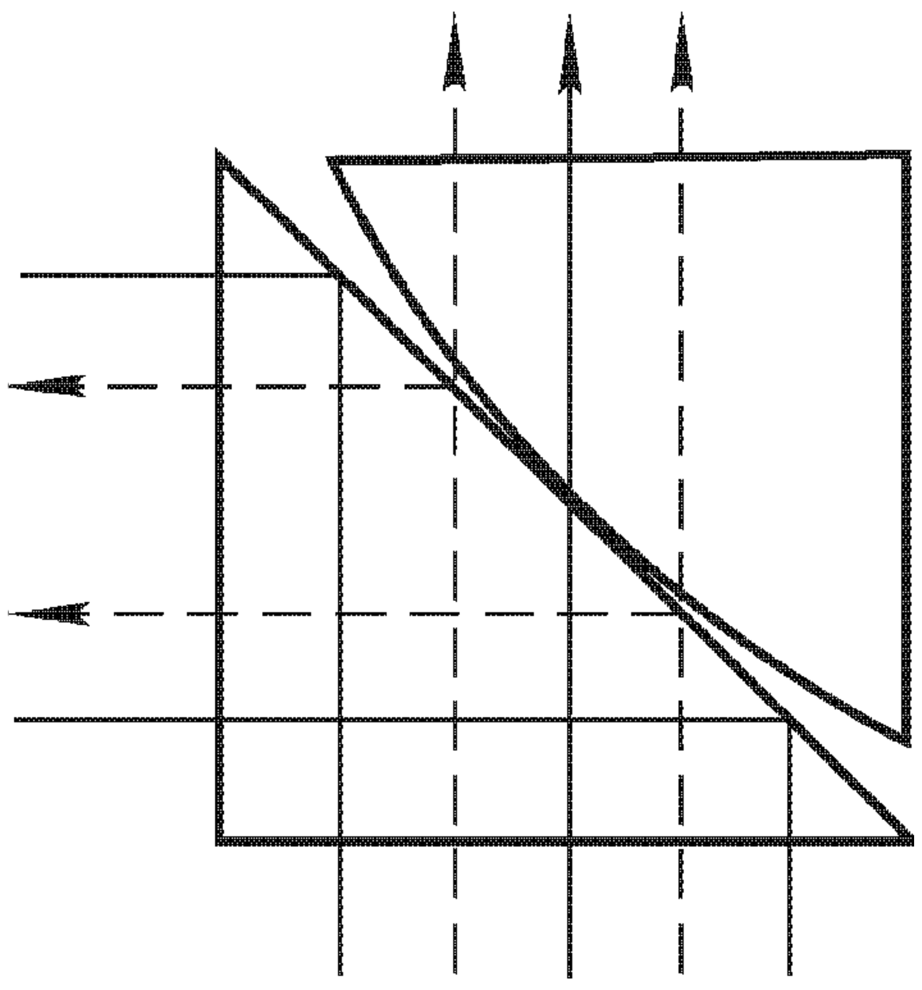


Рис. 242

Меняя толщину воздушного зазора, можно изменять интенсивность проходящего света. На этом основано устройство *светового телефона*, запатентованного фирмой Цейсс. Роль одной из сред играет прозрачная мембрана, колеблющаяся под действием падающего на нее звука. Свет, прошедший через зазор, меняет интенсивность в такт с изменениями силы звука. Падая на фотоэлемент, он возбуждает переменный ток, сила которого изменяется также в соответствии с изменениями силы звука. Этот ток усиливается и используется в дальнейшем.

Явления прохождения волн через тонкие зазоры, аналогичные описанному, не специфичны для оптики. Они имеют место для волн *любой природы* и могут происходить, когда фазовая скорость в зазоре превосходит фазовую скорость в окружающих средах. Легче всего они реализуются для длинных волн (радиотехника, акустика). Особо важное значение эти явления приобрели в ядерной и атомной физике, а также в теории твердого тела. Об этом будет говориться в пятом томе нашего курса.

3. При обыкновенном отражении оба отношения $R_\perp/\mathcal{E}_\perp$ и $R_\parallel/\mathcal{E}_\parallel$ всегда вещественны. При полном отражении они, вообще говоря, *комплексны*. Это значит, что при полном отражении *фаза волны испытывает скачок*, и притом, как правило, отличный от нуля или π. Пусть волна поляризована перпендикулярно к плоскости падения. Можно написать

$$\cos\varphi + i\sqrt{\sin^2\varphi - n^2} = Ae^{i\delta_\perp/2},$$

где A и $\delta_\perp$ — величины вещественные. Тогда

$$\cos\varphi - i\sqrt{\sin^2\varphi - n^2} = Ae^{-i\delta_\perp/2}, \qquad \frac{R_\perp}{\mathcal{E}_\perp} = e^{i\delta_\perp}.$$

Отсюда видно, что $\delta_\perp$ и есть интересующий нас скачок фазы. Приравнивая вещественные и мнимые части, получим

$$A\cos\frac{\delta_\perp}{2} = \cos\varphi, \qquad A\sin\frac{\delta_\perp}{2} = \sqrt{\sin^2\varphi - n^2}.$$

Отсюда

$$\operatorname{tg}\frac{\delta_\perp}{2} = \frac{\sqrt{\sin^2\varphi - n^2}}{\cos\varphi}. \tag{66.3}$$

Аналогично, для волны, поляризованной в плоскости падения,

$$\operatorname{tg}\frac{\delta_\parallel}{2} = \frac{\sqrt{\sin^2\varphi - n^2}}{n^2\cos\varphi}. \tag{66.4}$$

Скачки фаз $\delta_\perp$ и $\delta_\parallel$ не одинаковы. Если падающая волна поляризована линейно, то в отраженной волне между $\mathbf{E}_\parallel$ и $\mathbf{E}_\perp$ возникнет определенная разность фаз. Поэтому *отраженная волна будет, вообще говоря, поляризована эллиптически*. Допустим, что между $\mathbf{E}_\parallel$ и $\mathbf{E}_\perp$ в падающей волне нет разности фаз. Тогда разность фаз между ними в отраженной волне будет $\delta = \delta_\parallel - \delta_\perp$. С помощью формул (66.3) и (66.4) найдем

$$\operatorname{tg}\frac{\delta}{2} = \frac{\cos\varphi\sqrt{\sin^2\varphi - n^2}}{\sin^2\varphi} \tag{66.5}$$

Отсюда видно, что $\operatorname{tg}(\delta/2) > 0$, т.е. δ лежит в пределах от нуля до π. Значит, колебание $\mathbf{E}_\parallel$ опережает по фазе колебание $\mathbf{E}_\perp$, и эллиптическая поляризация отраженного света будет *левой*. Наоборот, если разность фаз между такими же колебаниями в падающей волне равна π, то эллиптическая поляризация отраженной волны будет *правой*.

Величина δ обращается в нуль при $\varphi = \varphi_0$ и $\varphi = \pi/2$. Между этими пределами она должна достигать максимума. Максимум достигается при $\cos\varphi = \sqrt{(1-n^2)/(1+n^2)}$, максимальная разность фаз δ_m определяется уравнением

$$\operatorname{tg}\frac{\delta_m}{2} = \frac{1-n^2}{2n}. \tag{66.6}$$

Для получения круговой поляризации отраженного света необходимо выполнение двух условий: 1) $\mathcal{E}_{\parallel} = \pm\mathcal{E}_{\perp}$; 2) $\delta = \pi/2$ (см. § 62, п. 4). Чтобы получить $\delta_m = \pi/2$, нужен показатель преломления, удовлетворяющий уравнению $1 = (1-n^2)/2n$. Отсюда $n = \sqrt{2} - 1 = 0{,}414$. Показатель преломления оптически более плотной среды относительно менее плотной будет $n' = 1/n = 2{,}41$. Если n' не меньше этого значения, то при однократном отражении можно получить разность фаз $\delta = 90^\circ$. В этом случае при $\mathcal{E}_{\perp} = \mathcal{E}_{\parallel}$ поляризация отраженного света будет *круговая* и притом левая. В видимой области спектра такой большой показатель преломления имеет только алмаз ($n' = 2{,}42$). Для всех прочих сред $n' < 2{,}41$. Поэтому для видимого света получить разность фаз $\delta = 90^\circ$ при однократном отражении невозможно. Но это возможно для коротких радиоволн. Для воды, например, $n' = \sqrt{\varepsilon_0} = 9$, $n = 1/9$; круговая поляризация, согласно (66.6), должна получиться при $\varphi = 6^\circ 29'$ и $\varphi = 44^\circ 38'$. Опыт подтвердил это предсказание.

Для стекла ($n' = 1{,}51$) максимальная разность фаз достигается при $\varphi = 51^\circ 20'$ и составляет $45^\circ 36'$. Разность фаз δ в точности равна 45° при $\varphi = 48^\circ 37'$ и при $\varphi = 54^\circ 37'$. Если применить двукратное отражение под одним из этих углов, то возникнет относительная разность фаз 90°. Этим воспользовался Френель для получения круговой поляризации. Он изготовил стеклянный параллелепипед с углом $A = 54^\circ 37'$ (рис. 243). Луч SM, падающий перпендикулярно на грань AB, претерпевает двукратное полное отражение на гранях AD и BC, после чего выходит из параллелепипеда в направлении NQ. При каждом из отражений колебание $\mathbf{E}_{\parallel}$ опережает по фазе колебание $\mathbf{E}_{\perp}$ на 45°. Если па-

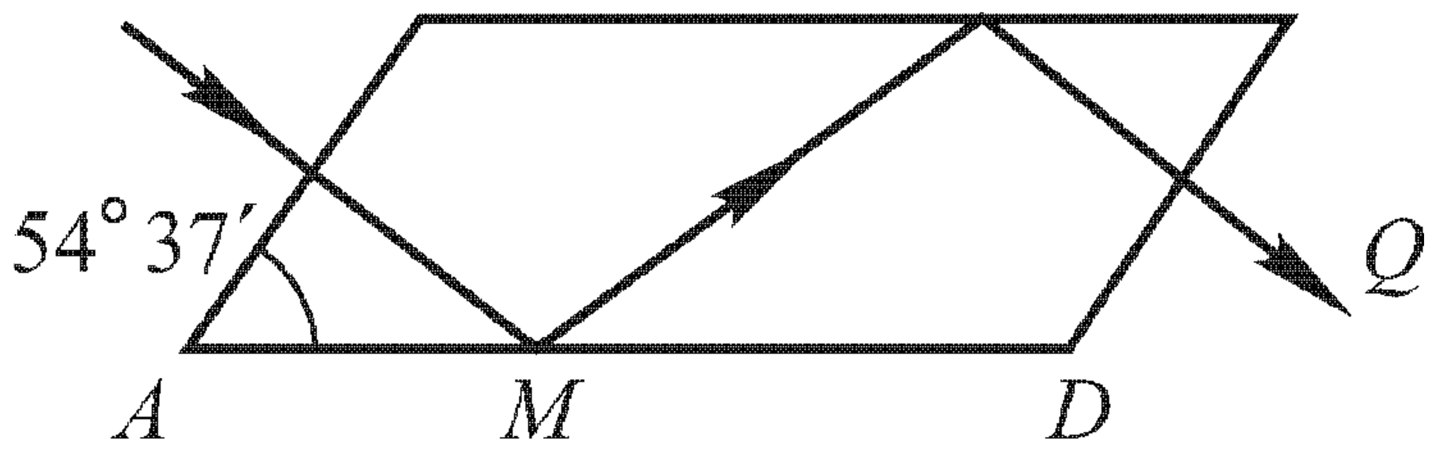


Рис. 243

дающий свет поляризован линейно и $\mathcal{E}_{\|} = \mathcal{E}_{\perp}$, то выходящий свет будет поляризован по левому кругу. Если его заставить вторично пройти через параллелепипед Френеля, то он снова превратится в линейно поляризованный, но плоскость поляризации повернется на 90° относительно своего исходного положения.

4. Возникновение скачка фазы при полном отражении можно понять, не обращаясь к формулам Френеля. Допустим ради определенности, что электрический вектор лежит в плоскости падения. Если угол падения строго равен предельному углу полного отражения φ_0, то волна во второй среде будет еще однородной. Она должна распространяться параллельно границе раздела. Так как однородные волны поперечны, то отсюда следует, что электрический вектор во второй, а следовательно, в силу граничных условий, и в первой среде нормален к границе раздела. При том выборе положительного направления электрического поля, которого мы придерживаемся (рис. 244), это означает, что фазы падающей и отраженной волн на границе раздела совпадают.

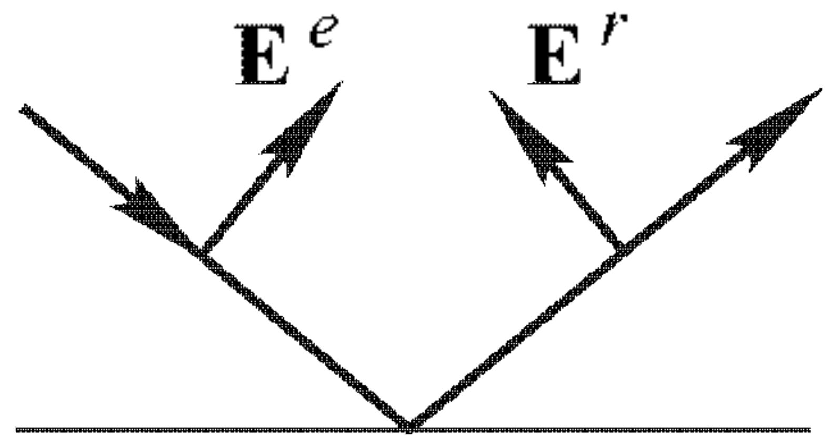


Рис. 244

Напротив, при $\varphi = 90°$ полное отражение сопровождается изменением фазы на 180°. В этом случае падающая и отраженная волны распространяются в одном и том же направлении — параллельно границе раздела. Граничные условия будут удовлетворены, если положить $R_{\|} = -\mathcal{E}_{\|}$. Падающая и отраженная волны гасят друг друга, в результате чего как в первой, так и во второй средах не будет никакого светового поля, и граничные условия выполняются тривиально. Но равенство $R_{\|} = -\mathcal{E}_{\|}$ означает, что при отражении появляется скачок фазы в 180°.

Следует заметить, что говорить об отражении света, когда угол падения точно равен 90°, не имеет смысла. Приведенное рассуждение показывает только, к какому пределу стремится амплитуда $R_{\|}$, когда угол падения приближается к 90°. Если отражение полное, то в первой среде образуются стоячие волны с узловыми плоскостями, параллельными границе раздела. По мере приближения угла падения к 90° расстояние между узловыми плоскостями увеличивается и стремится к бесконечности. При этом ближайшая узловая плоскость в пределе сливается с границей раздела. Одновременно исчезает поле во второй (оптически менее плотной) среде.

Итак, при $\varphi = \varphi_0$ скачок фазы при отражении равен нулю, а при $\varphi = 90°$ он составляет 180°. При этом во всем интервале изменения угла φ от φ_0 до 90° отражение должно оставаться полным (§ 64). Изобразим колебание электрического поля в отраженной волне на векторной диаграмме. При $\varphi = \varphi_0$ амплитуда изобразится вектором $\mathbf{A}(\varphi_0)$ (рис. 245), а само колебание — его проекцией на ось X (если заставить вектор $\mathbf{A}(\varphi_0)$ равномерно вращаться вокруг O с угловой скоростью ω). При $\varphi = \pi/2$ амплитуда изобразится вектором $\mathbf{A}(\pi/2)$, равным вектору $\mathbf{A}(\varphi_0)$, но противоположно направленным. Для промежуточного угла падения φ амплитуду, представляющую колебание, обозначим через $\mathbf{A}(\varphi)$. При возрастании φ от φ_0 до $\pi/2$ длина вектора $\mathbf{A}(\varphi)$ не может изменяться, так как отражение полное. С другой стороны, вектор $\mathbf{A}(\varphi)$, изменяясь непрерывно, должен из исходного положения $\mathbf{A}(\varphi_0)$ перейти в конечное положение $\mathbf{A}(\pi/2)$. Это можно осуществить *только поворотом его вокруг точки* O. Значит, если угол падения φ лежит между φ_0 и 90°, то вектор $\mathbf{A}(\varphi)$ должен быть наклонен к оси X. Угол наклона $\delta_{\|}$ и есть скачок фазы при отражении.

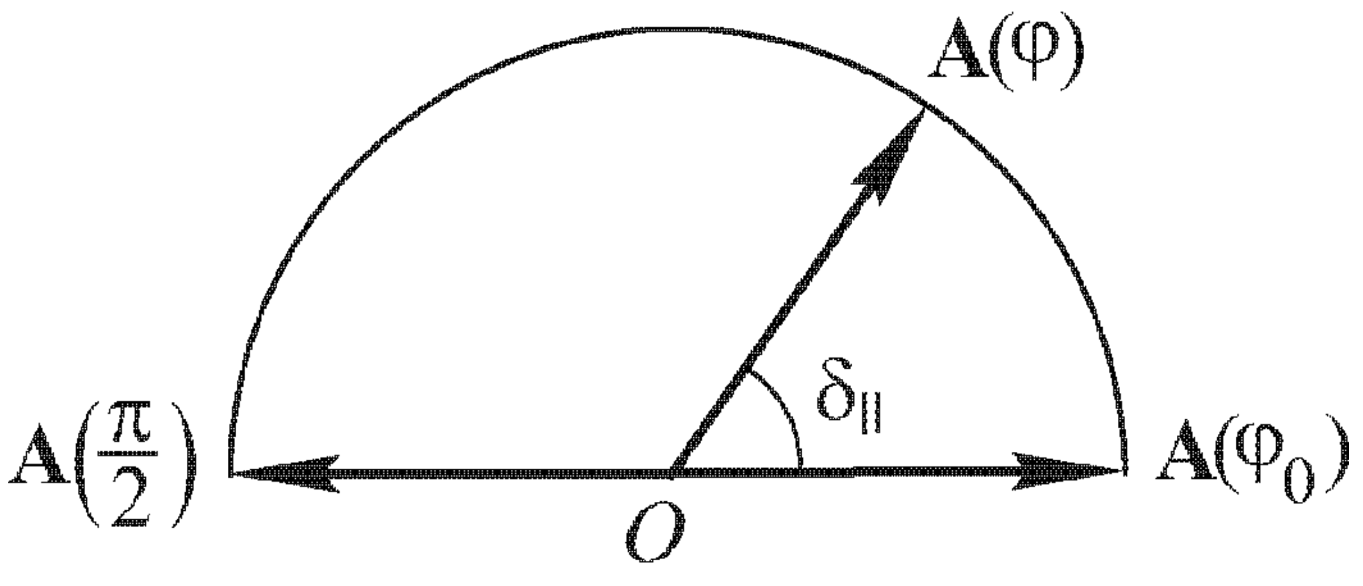


Рис. 245

5. Относительно перехода частичного отражения в полное сделаем следующее замечание. В области частичного отражения коэффициенты отражения $\rho_{\perp}$ и $\rho_{\|}$ меняются непрерывно и при $\varphi \to \varphi_0$ приближаются к единице. При $\varphi = \varphi_0$ они не испытывают разрыва. Однако их производные по φ претерпевают в этом месте разрыв. Вычислим производные по φ от коэффициентов Френеля $r_{\perp} = R_{\perp}/\mathcal{E}_{\perp}$ и $r_{\|} = R_{\|}/\mathcal{E}_{\|}$ в окрестности угла φ_0 в предположении, что $\varphi < \varphi_0$. Простое вычисление дает

$$\frac{dr_{\perp}}{d\varphi} = \frac{2}{\cos\psi}, \qquad \frac{dr_{\|}}{d\varphi} = \frac{2}{n^2\cos\psi}. \tag{66.7}$$

При $\varphi = \varphi_0$ $\cos\psi = 0$, и обе производные обращаются в бесконечность. Обращаются в бесконечность также $d\rho_{\perp}/d\varphi$ и $d\rho_{\|}/d\varphi$ $(\rho_{\perp} = r_{\perp}^2,\ \rho_{\|} = r_{\|}^2)$. Поэтому вблизи предельного угла φ_0 коэффициенты частичного отражения $\rho_{\perp}$ и $\rho_{\|}$ должны очень резко меняться с изменением угла падения. Этим объясняется происхождение *резкой границы* между областями частичного и полного отражений, которая бывает видна в рефрактометрах, основанных на полном отражении света.

§ 67. Прохождение света через плоскопараллельную пластинку. Просветление оптики

1. Допустим, что на поверхность отражающего тела нанесена пленка толщины l с показателем преломления n (рис. 246).

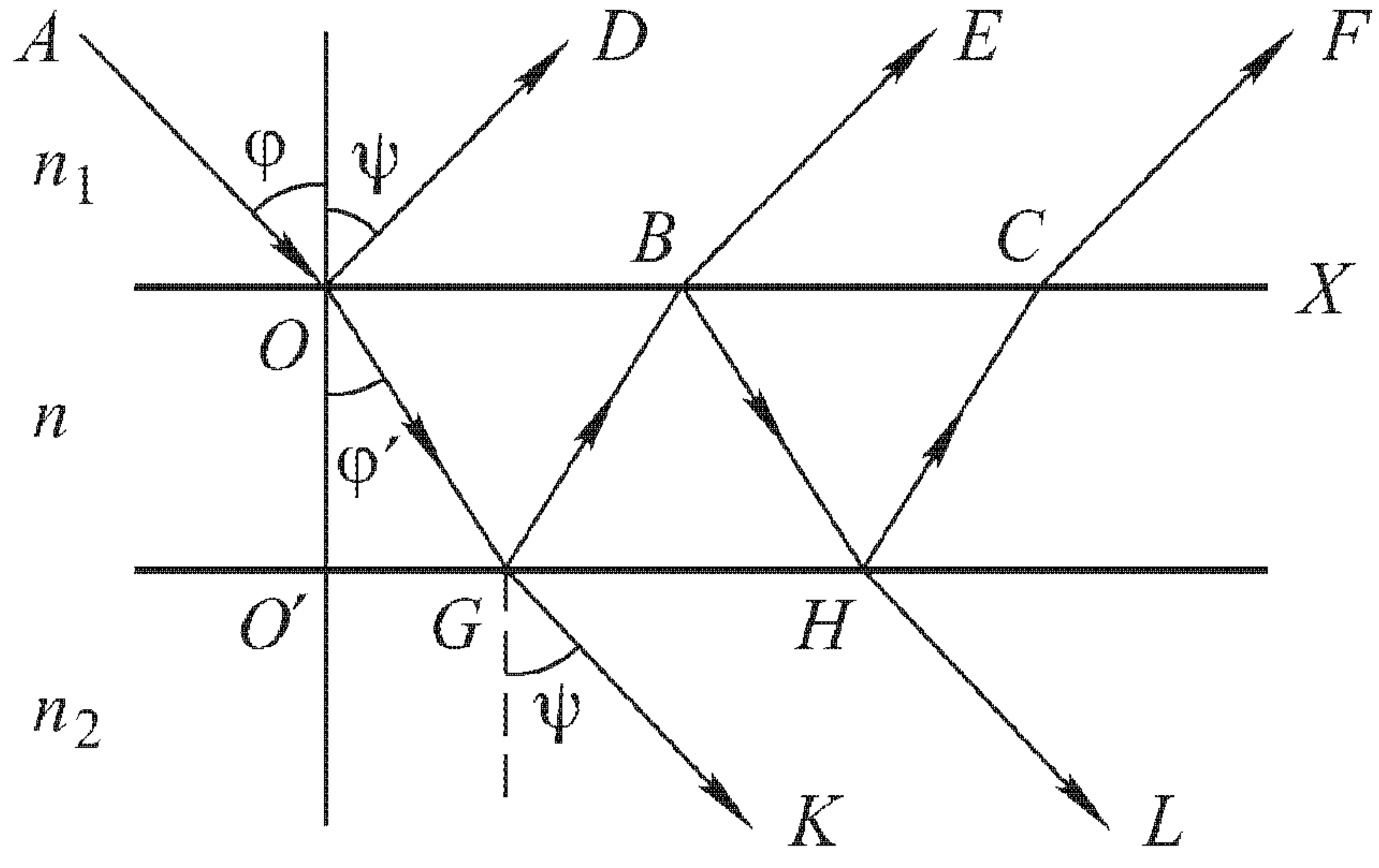


Рис. 246

Показатель преломления первой среды обозначим через n_1, а второй — через n_2. Пусть все показатели преломления n_1, n_2 и n постоянны (т.е. среды однородны), а падающая волна поляризована либо в плоскости падения, либо перпендикулярно к ней. Падающая волна

$$E^{(e)} = \mathcal{E}e^{i(\omega t - \mathbf{k}_1\mathbf{r})} \qquad (67.1)$$

(начало координат в O) претерпевает многократные отражения на границах пленки. В результате интерференции в первой среде возникает отраженная волна

$$E^{(r)} = Re^{i(\omega t - \mathbf{k}_1'\mathbf{r})}. \qquad (67.2)$$

Поле внутри пленки будет состоять из двух плоских волн:

$$E = Ae^{i(\omega t - \mathbf{k}\mathbf{r})}, \quad E' = A'e^{i(\omega t - \mathbf{k}'\mathbf{r})}. \qquad (67.3)$$

Наконец, во второй среде возникнет преломленная волна

$$E^{(d)} = De^{i(\omega t - \mathbf{k}_2\mathbf{r})}. \qquad (67.4)$$

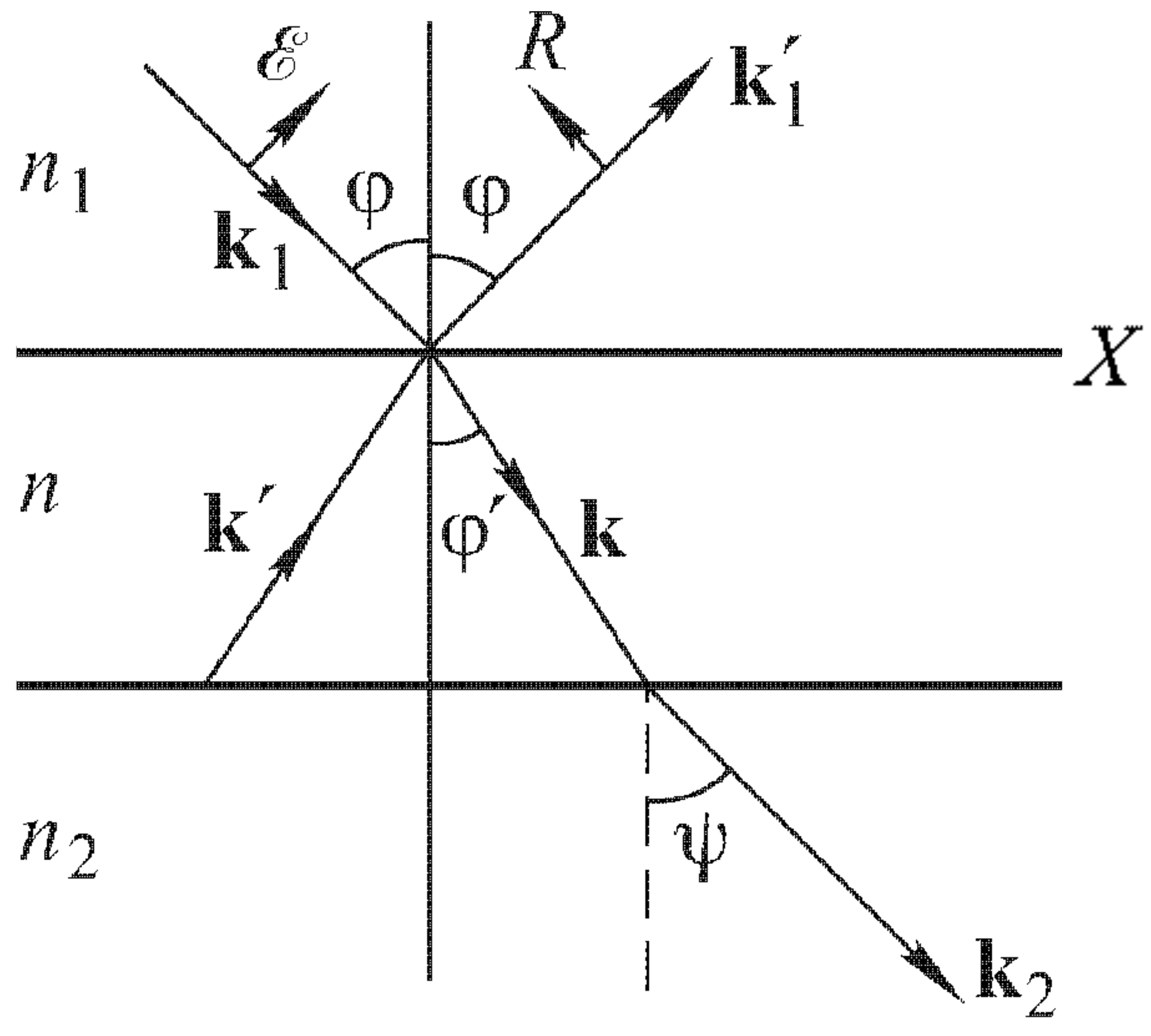


Рис. 247

Ввиду граничных условий, тангенциальные компоненты всех волновых векторов должны быть одинаковы. Этим однозначно определяются направления всех волн (см. § 64). Для определения неизвестных амплитуд R, A, A', D электромагнитная теория дает четыре независимых граничных условия (два на верхней поверхности пленки и два на нижней), из которых эти неизвестные могут быть найдены.

Однако мы предпочитаем решить задачу более коротким способом. На верхней границе сходятся четыре волны (рис. 247), из которых две приходящие, а две другие — уходящие. При написании соотношений между их комплексными амплитудами *можно отвлечься от наличия нижней границы*. Тогда формально можно рассуждать так, *как если бы многократных отражений не было*. Все приходящие волны можно рассматривать как падающие, в результате (однократных) отражений и преломлений которых возникают уходящие волны. Падающая волна с амплитудой $\mathcal{E}$, отражаясь от верхней границы, дает волну, идущую в направлении вектора $\mathbf{k}_1'$. Волна с амплитудой A', преломляясь на той же границе, дает волну, идущую в том же направлении. Результатом интерференции этих волн и является отраженная волна с амплитудой R. Это дает

$$R = r_1\mathcal{E} + d_1'A'. \tag{67.5}$$

Аналогично,

$$A = d_1\mathcal{E} + r_1'A'. \tag{67.6}$$

Здесь r_1 и d_1 означают коэффициенты Френеля на верхней границе пленки, а r_2 и d_2 — такие же коэффициенты на ее нижней границе (предполагается, что свет пересекает эти границы сверху вниз). Коэффициенты Френеля при обратном ходе лучей обозначаются теми же буквами, но со штрихами. С помощью формул Френеля нетрудно проверить, что эти величины удовлетворяют соотношениям

$$r' = -r, \quad r^2 + dd' = 1. \tag{67.7}$$

Такое же рассуждение можно провести и для нижней границы пленки. Надо только перенести начало координат в точку O', находящуюся на нижней поверхности пленки (см. рис. 246). Тогда волна (67.1) запишется в виде

$$E^{(e)} = \mathcal{E}e^{-ik_{1z}l}e^{i(\omega t - \mathbf{k}_1\mathbf{r}')},$$

откуда видно, что роль комплексной амплитуды будет играть величина $\mathcal{E}' = \mathcal{E}e^{-ik_{1z}l}$. Аналогично преобразуются комплексные амплитуды и остальных волн. В результате получим

$$A'E^{-ik_z'l} = r_2Ae^{-ik_zl}, \quad De^{-ik_{2z}l} = d_2Ae^{-ik_zl}. \tag{67.8}$$

Из полученных уравнений найдем

$$\begin{aligned} \frac{R}{\mathcal{E}} &= \frac{r_1 + r_2\exp(-2ik_0nl\cos\varphi)}{1 + r_1r_2\exp(-2ik_0nl\cos\varphi')}, \\ \frac{D}{\mathcal{E}} &= \frac{d_1d_2\exp[ik_0l(n_2\cos\psi - n\cos\varphi)]}{1 + r_1r_2\exp(-2ik_0nl\cos\varphi')}, \end{aligned} \tag{67.9}$$

где k_0 — волновое число в вакууме, а φ' и ψ — углы преломления на первой и второй границах. Для наглядности при выводе мы ссылались на чертеж, относящийся к однородным волнам. Но, очевидно, результаты (67.9) справедливы и в том случае, когда волны в пленке или во второй среде неоднородны.

2. При нормальном падении первая формула (67.9) переходит в

$$\frac{R}{\mathcal{E}} = \frac{r_1 + r_2e^{-2ik_0nl}}{1 + r_1r_2e^{-2ik_0nl}}. \tag{67.10}$$

Это выражение обращается в нуль, если выполняются два условия:

$$r_1 + r_2 \cos(2nk_0 l) = 0, \quad \sin(2nk_0 l) = 0.$$

Из второго условия получаем: $2nk_0 l = N\pi$, или

$$l = N\frac{\lambda_0}{4n} = N\frac{\lambda}{4}, \tag{67.11}$$

где N — целое число, а λ — длина волны в пленке. Теперь первое условие можно записать в виде: $r_1 + (-1)^N r_2 = 0$. Если N — нечетное, то $r_1 = r_2$; если же N четное, то $r_1 = -r_2$. Однако последнее равенство не может быть выполнено. Действительно,

$$r_1 = \frac{n - n_1}{n + n_1}, \quad r_2 = \frac{n_2 - n}{n_2 + n}.$$

Если бы $r_1 = -r_2$, то мы получили бы $n(n_2 - n_1) = 0$, откуда либо $n = 0$, что невозможно, либо $n_2 = n_1$, что не представляет интереса. Это показывает, что число N должно быть нечетным, а следовательно, $r_1 = r_2$. Отсюда

$$n = \sqrt{n_1 n_2}. \tag{67.12}$$

Таким образом, если $n = \sqrt{n_1 n_2}$, а толщина пленки равна $N\lambda/4$, где N — нечетное число, то отражательная способность обращается в нуль.

На этом результате основан один из методов увеличения поверхностной прозрачности стекол, применяемый в оптической промышленности (так называемое *просветление оптики*). Для стекла ($n = 1{,}5$) отражательная способность равна $R = (n-1)^2/(n+1)^2 = 0{,}04 = 4\,\%$, т.е. совсем невелика. Однако оптические приборы состоят из многих деталей, изготовленных из стекла. Отражение на границах их соприкосновения является главной причиной ослабления света при его прохождении через оптический прибор. Так, например, потери света в призменном бинокле составляют свыше 50 %, причем они почти целиком происходят за счет отражения света. Значительная доля отраженного света, благодаря последующим отражениям, доходит до глаза наблюдателя и, будучи в лучшем случае равномерно рассеянной, дает освещенный фон, ослабляющий контраст света и тени в изображении. Особенно вреден этот рассеянный свет в фотографических приборах. В лучшем случае он создает общую вуаль на эмульсии. При некоторых же неблагоприятных расположениях источников света могут получиться блики и дополнительные изображения.

Для увеличения поверхностной прозрачности стекла на его поверхности создается пленка с показателем преломления $n = \sqrt{n_1 n_2}$ и толщиной $\lambda/4$. Применение более толстых пленок (соответствующих $N = 3, 5$ и т.д.) нецелесообразно, так как условие полного исчезновения отражения может быть точно выполнено только для одной определенной длины волны и одного определенного угла падения. При $N = 1$ возрастание коэффициента отражения с изменением длины волны, а также с изменением угла падения получается наиболее медленным. Поэтому в этом случае можно добиться почти полного исчезновения отражения для сравнительно большого участка спектра и сравнительно широкого интервала углов падения.

Просветляющие пленки создаются либо путем выщелачивания из поверхности стекла его компонентов, либо, что лучше, путем напыления на поверхность стекла слоя посторонних веществ. При этом возникают трудности, связанные с тем, что показатель преломления пленки должен быть значительно меньше показателя преломления стекла. Например, если $n_2 = 1{,}52$, $n_1 = 1$, то $n = 1{,}23$. Твердых веществ с таким малым показателем преломления в природе не встречается. Поэтому пленку приходится делать пористой, причем, во избежание заметного рассеяния света, размеры пор должны быть весьма малы по сравнению с длиной волны. Но пористая пленка не обладает достаточной механической прочностью. Трудности могут быть преодолены путем применения двуслойных покрытий. Сначала просветляемая поверхность покрывается пленкой, показатель преломления которой значительно превосходит показатель преломления стекла, а затем пленкой с меньшим показателем преломления.

3. Условия отсутствия отражения (67.11) и (67.12) могут быть также получены из простых интерференционных соображений.

При малых коэффициентах отражения можно пренебречь волнами, претерпевшими многократные отражения на границах пленки. Тогда останутся две отраженные волны, из которых одна отразилась от передней, а другая от задней поверхности пленки. Чтобы эти волны гасили друг друга, должны соблюдаться два условия: 1) фазы их должны быть противоположны; 2) интенсивности их должны быть равны. Первое условие при нормальном падении приводит к соотношению (67.11). Второе условие, если пренебречь ослаблением волны за счет отражения от передней поверхности пленки, сводится к $r_1 = r_2$. Отсюда, как показано выше, получается формула (67.12). Недостаток приведенного доказательства в том, что оно не учитывает многократные отражения.

ЗАДАЧИ

1. Вывести формулы для отражения и преломления электромагнитных волн в предположении, что магнитная проницаемость отлична от единицы.

Ответ.

$$\frac{R_\perp}{\mathcal{E}_\perp}=\frac{\sqrt{\varepsilon_1/\mu_1}\,\cos\varphi-\sqrt{\varepsilon_2/\mu_2}\,\cos\psi}{\sqrt{\varepsilon_1/\mu_1}\,\cos\varphi+\sqrt{\varepsilon_2/\mu_2}\,\cos\psi},\quad \frac{D_\perp}{\mathcal{E}_\perp}=\frac{2\sqrt{\varepsilon_1/\mu_1}\,\cos\varphi}{\sqrt{\varepsilon_1/\mu_1}\,\cos\varphi+\sqrt{\varepsilon_2/\mu_2}\,\cos\psi},$$

$$\frac{R_\parallel}{\mathcal{E}_\parallel}=\frac{\sqrt{\varepsilon_2/\mu_2}\,\cos\varphi-\sqrt{\varepsilon_1/\mu_1}\,\cos\psi}{\sqrt{\varepsilon_2/\mu_2}\,\cos\varphi+\sqrt{\varepsilon_1/\mu_1}\,\cos\psi},\quad \frac{D_\parallel}{\mathcal{E}_\parallel}=\frac{2\sqrt{\varepsilon_1/\mu_1}\,\cos\varphi}{\sqrt{\varepsilon_2/\mu_2}\,\cos\varphi+\sqrt{\varepsilon_1/\mu_1}\,\cos\psi}. \tag{67.13}$$

2. Какой вид принимают формулы (67.13), если $\varepsilon_1=\varepsilon_2=1$?

Ответ.

$$\begin{aligned}\frac{R_\perp}{\mathcal{E}_\perp}&=\frac{n_2\cos\varphi-n_1\cos\psi}{n_2\cos\varphi+n_1\cos\psi}, & \frac{D_\perp}{\mathcal{E}_\perp}&=\frac{2n_2\cos\varphi}{n_2\cos\varphi+n_1\cos\psi},\\ \frac{R_\parallel}{\mathcal{E}_\parallel}&=\frac{n_1\cos\varphi-n_2\cos\psi}{n_1\cos\varphi+n_2\cos\psi}, & \frac{D_\parallel}{\mathcal{E}_\parallel}&=\frac{2n_2\cos\varphi}{n_1\cos\varphi+n_2\cos\psi}.\end{aligned} \tag{67.14}$$

Формулы для отражения в точности совпадают с формулами Френеля, если только поменять местами $\perp$ и $\parallel$-составляющие. Этот факт был отмечен Гельмгольцем.

3. Показать, что отражательная способность среды для радиоволн обращается в нуль, если $\varepsilon=\mu$.

4. Останется ли справедливым закон Брюстера для радиоволн, если магнитные проницаемости сред μ_1 и μ_2 отличны от единицы?

Ответ. Если закон Брюстера имеет место, то угол Брюстера, при котором не отражается составляющая $E_\parallel$ электрического поля, определяется выражением

$$\operatorname{tg}\varphi_B=\sqrt{\frac{\varepsilon_2}{\varepsilon_1}\,\frac{\varepsilon_2\mu_1-\varepsilon_1\mu_2}{\varepsilon_2\mu_2-\varepsilon_1\mu_1}}. \tag{67.15}$$

Возможен случай, когда не будет отражаться составляющая $E_\perp$. Угол, при котором это имеет место, находится из уравнения

$$\operatorname{tg}\varphi'_B=\sqrt{\frac{\mu_2}{\mu_1}\,\frac{\varepsilon_2\mu_1-\varepsilon_1\mu_2}{\varepsilon_1\mu_1-\varepsilon_2\mu_2}}. \tag{67.16}$$

Оба случая взаимно исключают друг друга, так как знаки подкоренных выражений в (67.15) и (67.16) противоположны. (Предполагается, что ε и μ существенно положительны.) Таким образом, всегда существует угол, при котором не отражается либо составляющая $E_\parallel$, либо составляющая $E_\perp$.

5. Проверить, что коэффициенты Френеля удовлетворяют соотношениям (67.7).

6. Пользуясь формулами Френеля, показать, что линейно поляризованный свет остается линейно поляризованным после отражения на границе раздела двух прозрачных изотропных сред во всех случаях, за исключением случаев полного отражения.

7. Угол между плоскостью колебаний поляризованного света и плоскостью падения называется *азимутом колебания*. Найти азимут преломленной волны γ и азимут отраженной волны β, если азимут падающей волны α, а угол падения φ.

Ответ. $\operatorname{tg}\gamma=\cos(\varphi-\psi)\operatorname{tg}\alpha$, $\operatorname{tg}\beta=-\dfrac{\cos(\varphi-\psi)}{\cos(\varphi+\psi)}\operatorname{tg}\alpha$.

8. 1) Найти угол Брюстера для света, отраженного от стекла с показателем преломления $n=1{,}5$; 2) найти для этого угла *степень поляризации* *преломленного света*, т.е. величину $\Delta = \dfrac{I_\perp - I_\parallel}{I_\perp + I_\parallel}$, где $I_\parallel$ и $I_\perp$ — интенсивности отраженных волн, поляризованных соответственно в плоскости падения и перпендикулярно к ней. Падающий свет — естественный.

Ответ. $\varphi_B = 56^\circ 19'$, $\Delta = \dfrac{4n^2 - (n^2+1)^2}{4n^2 + (n^2+1)^2} = -0{,}08$.

9. Естественный свет падает под углом Брюстера из воздуха на поверхность стекла с показателем преломления $n = 1{,}5$. Найти интенсивность I_r отраженного света, зная интенсивность падающего света I_e.

Ответ. $\dfrac{I_r}{I_e} = \dfrac{1}{2}\left(\dfrac{n^2-1}{n^2+1}\right)^2 = 0{,}74$.

10. Естественный свет падает под углом Брюстера из воздуха на поверхность диэлектрика с показателем преломления n. Найти амплитуду $D_\parallel$ преломленной волны, если амплитуда падающей волны той же поляризации равна $\mathcal{E}_\parallel$.

Ответ. $D_\parallel = \mathcal{E}_\parallel / n$.

11. На боковую грань призмы, изготовленной из стекла с показателем преломления $n = 1{,}5$, падает под углом Брюстера φ_B световой пучок, электрический вектор которого лежит в плоскости падения. Каким должен быть преломляющий угол A призмы, чтобы свет прошел через нее, не испытав потерь на отражение?

Ответ. $A = \pi - 2\varphi_B = 68^\circ$.

12. При каких условиях луч света, падающий на боковую грань прозрачной изотропной призмы с преломляющим углом $A = 60^\circ$, проходит через нее без потерь на отражение?

Ответ. Электрический вектор должен лежать в плоскости падения. Показатель преломления призмы должен быть равен

$$n = \operatorname{tg}\frac{A}{2} = \sqrt{3} = 1{,}73.$$

13. Свет падает из среды 1 на среду 2 под углом φ и преломляется под углом ψ. Доказать, что коэффициент отражения не изменится, если свет будет падать из среды 2 на среду 1 под углом ψ.

14. Пользуясь формулами Френеля, показать, что при отражении плоской электромагнитной волны от идеального зеркала, покрытого сверху слоем прозрачного диэлектрика, амплитуда отраженной волны равна амплитуде падающей. Изменяется лишь фаза волны, как это и должно быть согласно закону сохранения энергии.

Решение. При отражении от идеального зеркала может меняться только фаза, но не амплитуда волны. Поэтому коэффициент Френеля r_2 в формуле (67.10) должен иметь вид $r_2 = e^{i\alpha}$. Вводя еще обозначение $\delta = 2k_0 nl$, получим из той же формулы

$$\frac{R}{\mathcal{E}} = \frac{r_1 + e^{i(\alpha-\delta)}}{1 + r_1 e^{i(\alpha-\delta)}}.$$

Отсюда $|R| = |\mathcal{E}|$.

15. Свет, частично отражаясь, проходит через две параллельные полупрозрачные плоскости. Коэффициенты отражения и пропускания первой из них равны ρ_1 и b_1, а второй ρ_2 и b_2 соответственно. Степень монохроматичности падающего света невелика, так что интерференции не возникает, а имеет место сложение интенсивностей света. Найти коэффициенты отражения ρ и пропускания b для системы обеих плоскостей.

Решение.

$$\rho = \rho_1 + \rho_2 b_1^2 + \rho_2 b_1^2 \rho_1 \rho_2 + \ldots = \rho_1 + \frac{\rho_2 b_1^2}{1-\rho_1\rho_2},$$
$$b = b_1 b_2 + b_1 b_2 \rho_1 \rho_2 + b_1 b_2 (\rho_1\rho_2)^2 + \ldots = \frac{b_1 b_2}{1-\rho_1\rho_2}.$$

Используя соотношения $\rho_1 + b_1 = 1$, $\rho_2 + b_2 = 1$, полученные результаты можно записать в виде

$$\rho = \frac{\rho_1 + \rho_2 - 2\rho_1\rho_2}{1-\rho_1\rho_2}, \quad b = \frac{(1-\rho_1)(1-\rho_2)}{1-\rho_1\rho_2}. \tag{67.17}$$

При $\rho_1 = \rho_2$ отсюда получаем

$$\rho = \frac{2\rho_1}{1+\rho_1}, \quad b = \frac{1-\rho_1}{1+\rho_1}. \tag{67.18}$$

16. Свет, частично отражаясь, проходит через систему m параллельных полупрозрачных плоскостей. Коэффициенты отражения и пропускания каждой из них равны ρ и b. Найти коэффициент отражения ρ_m и коэффициент пропускания b_m всей системы m плоскостей. (Относительно падающего света см. предыдущую задачу.)

Решение. Присоединим к системе m плоскостей одну такую же $(m+1)$-ю плоскость. Первые m плоскостей можно заменить одной плоскостью с коэффициентом отражения ρ_m. Тогда задача будет сведена к предыдущей, и мы получим для коэффициентов отражения и пропускания $(m+1)$ плоскостей

$$\rho_{m+1} = \frac{\rho_m + \rho - 2\rho_m\rho}{1-\rho_m\rho},$$
$$b_{m+1} = \frac{(1-\rho_m)(1-\rho)}{1-\rho_m\rho}.$$

Отсюда методом доказательства от m к $m+1$ нетрудно получить

$$\rho_m = \frac{m\rho}{1+(m-1)\rho},$$
$$b_m = \frac{1-\rho}{1+(m-1)\rho}. \tag{67.19}$$

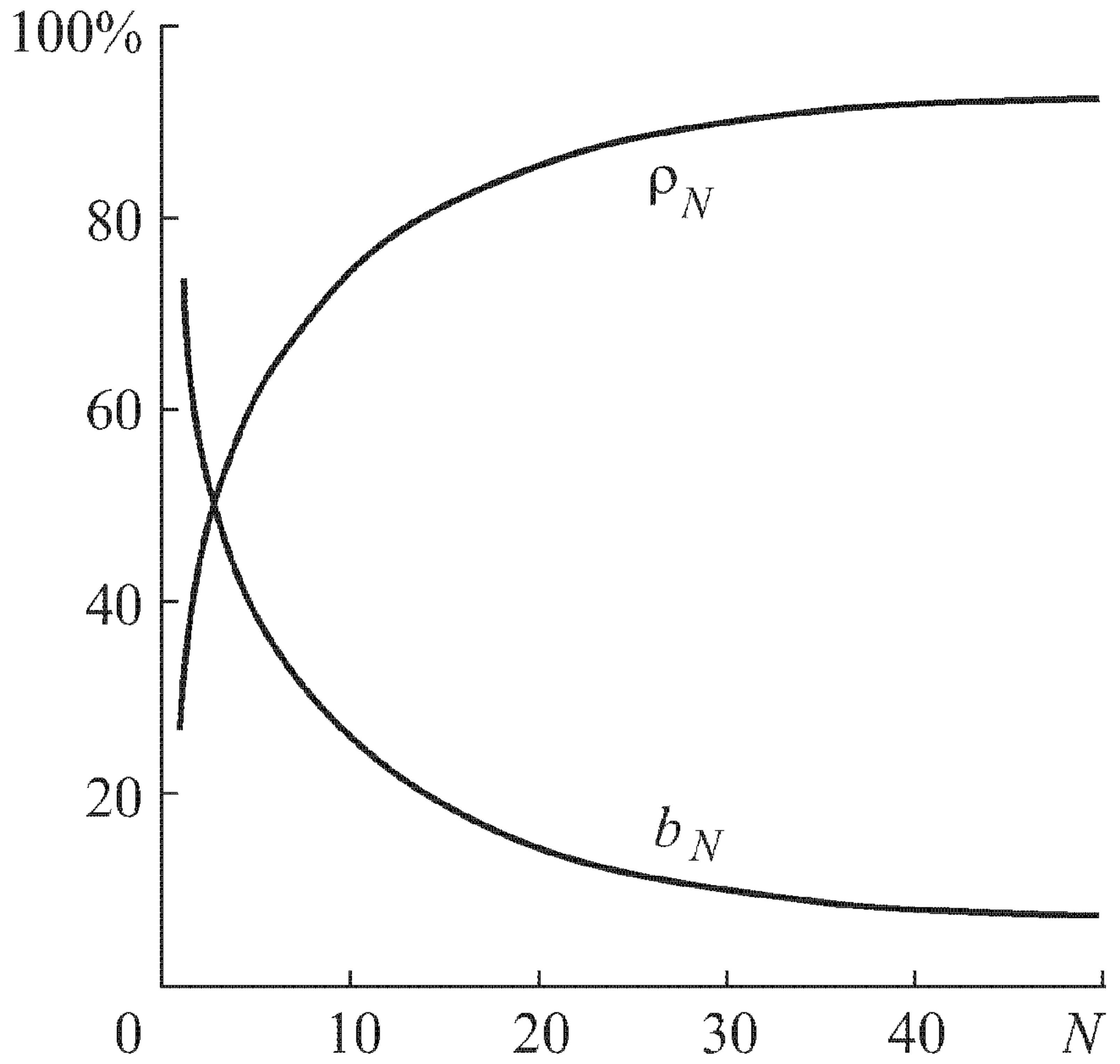


Рис. 248

17. Стопа Столетова (1839–1896) состоит из плоскопараллельных стеклянных пластинок с показателем преломления $n = 1{,}5$. На нее под углом Брюстера падает свет, поляризованный перпендикулярно к плоскости падения. Найти выражения и начертить график для коэффициентов отражения и пропускания стопы в зависимости от числа N пластинок.

Ответ. $\rho_N = \dfrac{2N}{2N+5{,}76}$, $b_N = \dfrac{5{,}76}{2N+5{,}76}$ (рис. 248).

18. Естественный свет падает под углом φ на стопу Столетова, состоящую из N стеклянных пластинок с показателем преломления n. Найти степени поляризации Δ и Δ' прошедшего и отраженного света, а также отношение интенсивностей $I_\perp/I_\parallel$ главных компонент прошедшего света.

Ответ.

$$\Delta = \frac{N(\rho_\perp - \rho_\parallel)}{(2N-1)\rho_\perp\rho_\parallel - (N-1)(\rho_\perp + \rho_\parallel) - 1},$$

$$\Delta' = \frac{\rho_\perp - \rho_\parallel}{2(N-1)\rho_\perp\rho_\parallel + (\rho_\perp + \rho_\parallel)}, \qquad \frac{I_\perp}{I_\parallel} = \frac{1-\rho_\perp}{1-\rho_\parallel}\,\frac{(2N-1)\rho_\parallel + 1}{(2N-1)\rho_\perp + 1},$$

где $\rho_\parallel$ и $\rho_\perp$ — коэффициенты отражения волн, поляризованных в плоскости падения и перпендикулярно к ней, для одной отражающей поверхности:

$$\rho_\perp = \left(\frac{\cos\varphi - n\cos\psi}{\cos\varphi + n\cos\psi}\right)^2, \qquad \rho_\parallel = \left(\frac{n\cos\varphi - \cos\psi}{n\cos\varphi + \cos\psi}\right)^2.$$

19. Падающий свет поляризован линейно с азимутом колебаний, равным $+45°$ [1]. Можно ли путем однократного отражения превратить его в свет, поляризованный по правому кругу?

Ответ. Нельзя.

20. Какой должен быть минимальный показатель преломления параллелепипеда Френеля, чтобы при азимуте колебаний падающего света в $+45°$ выходящий свет был поляризован по правому кругу?

Ответ. $n = \dfrac{1 + \sin(3\pi/8)}{\cos(3\pi/8)} = 5{,}028$. В оптическом диапазоне спектра этот случай осуществить нельзя. Его можно было бы осуществить с более длинными электромагнитными волнами.

21. Линейно поляризованная электромагнитная волна с азимутом колебаний, равным $+135°$, отражается на границе вода–воздух. Диэлектрическая проницаемость воды $\varepsilon = 81$. Под каким углом должна падать эта волна, чтобы отраженная волна получилась поляризованной по кругу? Какая при этом будет поляризация: правая или левая?

Ответ. $6°29'$ или $44°38'$. Правая.

22. Линейно поляризованный луч с азимутом колебаний $+135°$ падает перпендикулярно на грань AB стеклянной призмы $ABCD$ (рис. 249) и, испытав три раза полное отражение, выходит из нее. Каков должен быть преломляющий угол A призмы, чтобы вышедший свет был поляризован по кругу, если показатель преломления стекла призмы равен 1,52? Какая получится поляризация вышедшего света: правая или левая?

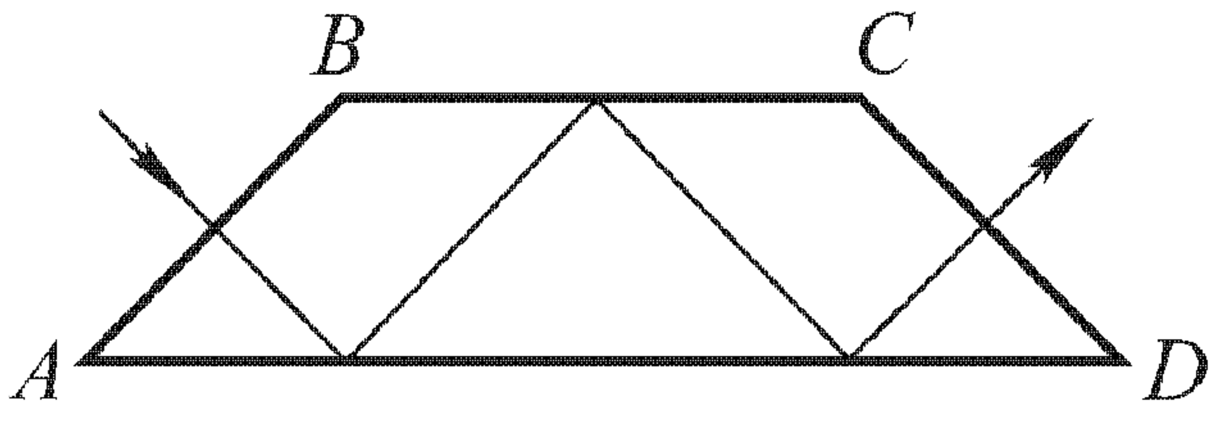


Рис. 249

Ответ. $69°21'$ или $42°46'$. Правая.

23. Каким должен быть минимальный показатель преломления призмы, описанной в предыдущей задаче, чтобы при азимуте колебаний падающего света, равном $+45°$, выходящий свет был поляризован по правому кругу? Какой при этом должен быть угол A?

---

[1] Азимут колебаний падающей волны может изменяться от $-\pi/2$ до $+\pi/2$. Он считается положительным, если $\mathcal{E}_\parallel/\mathcal{E}_\perp > 0$, и отрицательным, если $\mathcal{E}_\parallel/\mathcal{E}_\perp < 0$.

Ответ. $n = 1/(\sqrt{2} - 1) = 2{,}4143$; $A = 35°34'$.

24. С помощью векторной диаграммы показать, что скачок фазы при полном отражении превосходит вдвое скачок фазы, испытываемый преломленной (поверхностной) волной.

§ 68. Распространение света в среде с точки зрения молекулярной оптики

1. С точки зрения атомистических представлений всякую среду следует рассматривать как вакуум, в который вкраплены атомы вещества. (Молекулы можно рассматривать тоже как атомы.) Под действием падающей волны, а также излучений соседних атомов внутри каждого атома возбуждаются колебания электронов и атомных ядер. Вследствие этого атомы становятся источниками вторичных сферических волн, распространяющихся между этими частицами *со скоростью света в вакууме с*. Эти волны когерентны, поскольку они возбуждаются одной и той же падающей волной. Их интерференция между собой и с падающей волной определяет волновое поле во всем пространстве. В частности, отраженная волна возникает в результате интерференции вторичных волн, вышедших из среды в вакуум, с которым она граничит.

Почему же в среде свет распространяется с иной скоростью, чем в вакууме? Вопрос этот надо уточнить, указав, о какой скорости идет речь. В теории отражения и преломления света основной интерес представляет *фазовая скорость*, поскольку она определяет показатель преломления среды, а следовательно, и законы отражения и преломления волн на границе раздела сред. Отличие фазовой скорости света в среде от скорости света в вакууме вкратце объясняется тем, что в каждую точку пространства вторичные волны приходят не только от атомов, расположенных вдоль луча, проходящего через рассматриваемую точку, но и от множества других атомов, расположенных в стороне от него.

Более подробное рассмотрение приводится ниже для точечных атомов. В поле световой волны атомы приобретают меняющиеся во времени дипольные моменты и излучают как точечные диполи Герца. Для наших целей достаточно знать поле излучения такого диполя *в волновой зоне*. Оно определяется только составляющей дипольного момента $\mathbf{p}_\perp$, перпендикулярной к направлению излучения. Параллельная составляющая $\mathbf{p}_\parallel$ в волновой зоне на излучение диполя Герца не влияет.

2. Допустим, что в вакууме вдоль оси X распространяется плоская монохроматическая волна $\mathbf{E}_0 e^{i(\omega t - k_0 x)}$, на пути которой перпендикулярно к оси X поставлен бесконечно тонкий плоскопараллельный слой толщины $d\xi$, состоящий из точечных неподвижных атомов, равномерно распределенных по объему слоя (рис. 250). Следуя Рэлею, выясним влияние такого слоя на фазу колебаний в какой-то удаленной точке $A(x)$, расположенной впереди слоя. Дипольные моменты атомов слоя, возбужденные падающей волной, можно представить в виде $\mathbf{p} = \mathbf{p}_0 e^{i(\omega t - k_0\xi)}$, где ξ — абсцисса слоя. Предположим, что точка A расположена в волновой зоне ближайшего, а следовательно, и всех остальных диполей слоя. Тогда электрическое поле каждого диполя в точке A будет

$$\frac{k_0^2}{\mathbf{r}}\mathbf{p}_{0\perp}e^{i(\omega t - k_0\xi - k_0 r)}, \tag{68.1}$$

где r — расстояние от диполя (см. т. III, § 141). Такие выражения надо просуммировать по всем диполям слоя. Применим для этого метод кольцевых зон Френеля. Из теории таких зон известно, что результирующая напряженность $d\mathbf{E}_1$ всех диполей слоя в точке A будет равна половине напряженности поля, возбуждаемого в той же точке диполями одной только центральной зоны. Таким образом, для нахождения $d\mathbf{E}_1$ надо просуммировать выражение (68.1) по всем диполям центральной зоны и результат разделить на два. Вторичные волны, исходящие от края центральной зоны, отстают по фазе на π от волны, исходящей из ее центра O, а следовательно, и от падающей волны. Отставание по фазе вторичных волн, исходящих от остальных точек той же зоны, будет промежуточным. Таким образом, возникнет замедление скорости распространения фазы волны в результате прохождения ее через слой вещества.

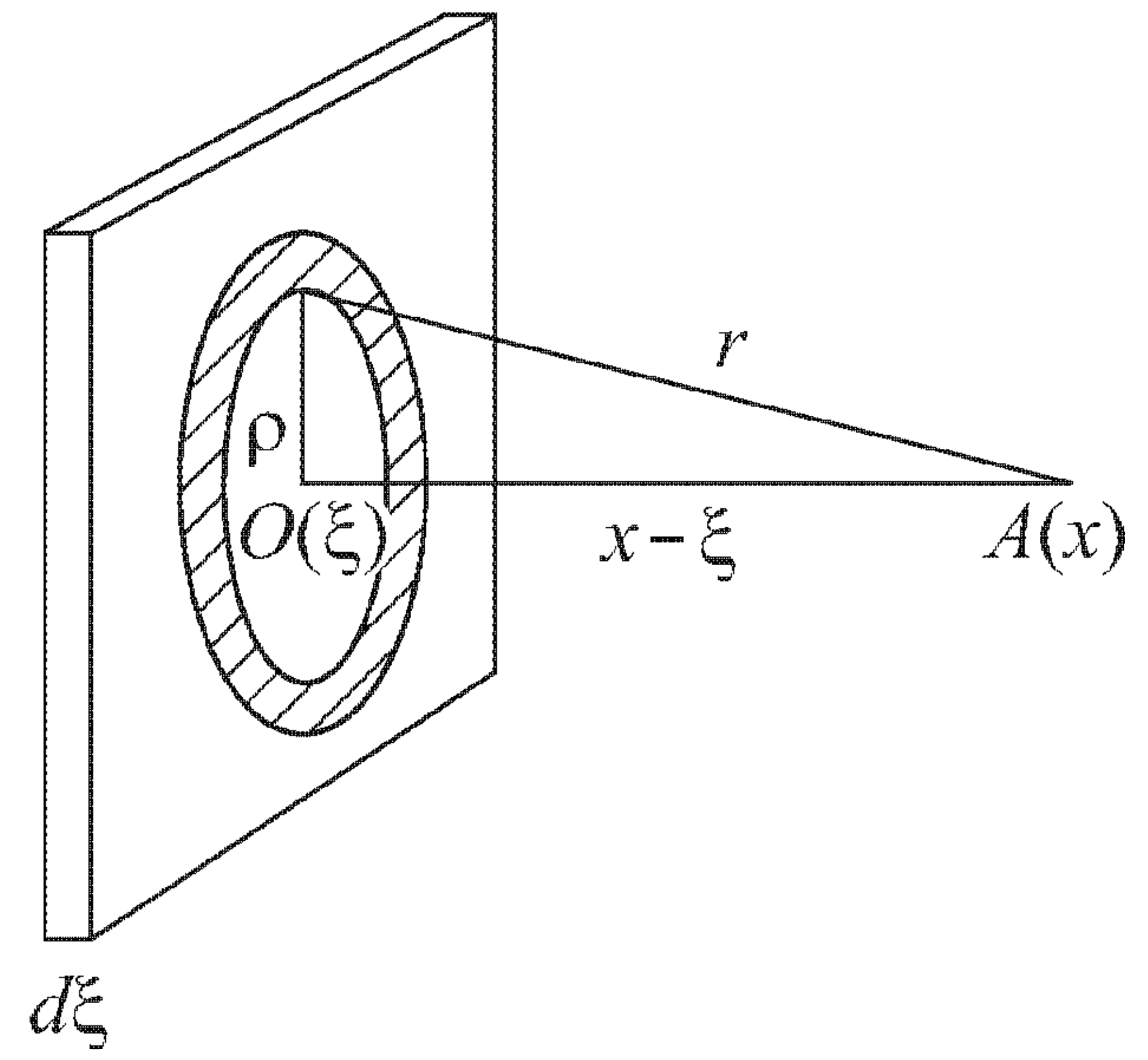


Рис. 250

Для фактического выполнения расчета заменим суммирование интегрированием. Возьмем кольцо с внутренним радиусом ρ и наружным $\rho + d\rho$, заштрихованное на рис. 250. В элементарном объеме $dV = 2\pi\rho\,d\rho\,d\xi$ находится $N\,dV$ диполей (N — число диполей в единице объема). Для возможности аппроксимации сумм интегралами и применимости метода зон Френеля предположим, что число $N\,dV$ еще достаточно велико. На это число надо умножить выражение (68.1), проинтегрировать по центральной зоне и результат разделить на два. Из соотношения $\rho^2 = r^2 - (x-\xi)^2$ при постоянном ξ получаем $\rho\,d\rho = r\,dr$ и вводим в качестве переменной интегрирования расстояние r. В пределах центральной зоны величину $\mathbf{p}_{0\perp}$ можно считать постоянной и равной $\mathbf{p}_0$. Тогда интегрирование сведется к

$$\int\limits_{x-\xi}^{x-\xi+\lambda/2} e^{-ik_0 r}\,dr = -\frac{2}{k_0}i,$$

а после введения коэффициента перед интегралом получится

$$d\mathbf{E}_1 = -i\cdot 2\pi k_0 N\mathbf{p}_0\,d\xi\,e^{i(\omega t-k_0 x)}.$$

Интегрированием по остальным зонам убеждаемся, что из-за убывания $\mathbf{p}_{0\perp}$ их действия медленно убывают с возрастанием номера зоны n и при $n\to\infty$ стремятся к нулю. Это может служить обоснованием применимости метода зон Френеля к рассматриваемому случаю.

Добавив $d\mathbf{E}_1$ к падающей волне, найдем полное поле в точке A:

$$\mathbf{E} = (\mathbf{E}_0 - i\cdot 2\pi k_0 N\mathbf{p}_0\,d\xi)e^{i(\omega t-k_0 x)} = \mathbf{E}_0 e^{i(\omega t-k_0 x-d\Phi)},$$

где введено обозначение

$$d\Phi = \frac{2\pi k_0 N p_0}{E_0}\,d\xi.$$

Таким образом, наличие слоя вносит дополнительное отставание по фазе $d\Phi$. Если толщина слоя l конечна, то отставание по фазе будет равно

$$\Phi = \frac{2\pi k_0 N p_0}{E_0}l. \tag{68.2}$$

В этой формуле содержится *принципиальное объяснение* замедления фазовой скорости световой волны при ее распространении в веществе.

Для завершения расчета надо было бы найти связь между амплитудами $\mathbf{p}_0$ и $\mathbf{E}_0$. В общем случае это весьма сложная задача, так как дипольный момент атома $\mathbf{p}$ определяется не средним макроскопическим полем $\mathbf{E}$, а *микроскопическим полем, действующим на атомы среды.* Только для не очень плотных газов (когда $n-1\ll 1$) оба поля практически совпадают. Тогда $\mathbf{p}_0 = \beta\mathbf{E}_0$, где β — поляризуемость атома, связанная с диэлектрической проницаемостью ε и показателем преломления n соотношением

$$\varepsilon = n^2 = 1 + 4\pi N\beta = 1 + 4\pi N p_0/E_0.$$

Используя это соотношение, из (68.2) найдем

$$\Phi = \frac{n^2-1}{2}\,k_0 l \approx (n-1)k_0 l,$$

что совпадает с результатом феноменологического рассмотрения.

3. В феноменологической теории показатель преломления вводится с помощью *макроскопических уравнений Максвелла.* Последние предполагают, что в каждом элементарном объеме, линейные размеры которого малы по сравнению с длиной волны, содержится еще *очень много атомов.* Молекулярное рассмотрение, приведенное выше, показывает, что *это условие не обязательно.* Показатель преломления можно определить *через сдвиг фазы*, который вносит вещество, стоящее на пути световой волны. Такой сдвиг был вычислен выше в предположении, что велико число атомов во всяком элементе объема порядка $dV = 2\pi\rho\, d\rho\, d\xi$. А этому условию можно удовлетворить для сколь угодно разреженной среды, если только точку наблюдения A отодвинуть от слоя $d\xi$ достаточно далеко. Так, можно говорить о *показателе преломления рентгеновских лучей*, хотя макроскопические уравнения Максвелла на них не распространяются. Не лишено смысла говорить о *показателе преломления межпланетного и межзвездного пространства*, хотя плотность вещества в нем и ничтожна (не превышает примерно одного атома в кубическом сантиметре).

4. Если бы точку наблюдения A поместить перед слоем, то наше вычисление привело бы к волне, распространяющейся в противоположном направлении, т.е. к *отраженной волне.* Если средние межатомные расстояния меньше длины волны и атомы распределены в пространстве равномерно, то не возникнет никаких волн помимо прошедшей прямо и отраженной.

Не так будет, когда межатомные расстояния *больше длины волны.* Если атомы в среде распределены регулярно, например находятся в узлах кристаллической решетки, то вторичные волны, излучаемые атомами, когерентны, и будут складываться напряженности волновых полей. Условия интерференционного усиления вторичных волн могут выполняться не только в направлениях падающего и отраженного света, но и для некоторых других направлений. Возникнет дискретный ряд плоских волн, распространяющихся в различных направлениях (*интерференционное рассеяние*). Такой случай реализуется при дифракции коротких рентгеновских волн на кристаллической решетке. Если же атомы среды распределены в пространстве хаотически, то вторичные волны при рассмотрении бокового рассеяния ведут себя как некогерентные: складываются их интенсивности.

5. До сих пор не учитывалось *тепловое движение атомов.* Объясним теперь, как при наличии такового в среде может распространяться регулярная волна и как может возникнуть правильное отражение от зеркальных поверхностей твердых и жидких тел.

Рассмотрим сначала газы. Между столкновениями атомы газа движутся прямолинейно и равномерно. Из-за эффекта Доплера атомы, движущиеся с различными скоростями, излучают свет с различными частотами. Казалось бы, что никакой интерференции при таких условиях возникнуть не может. На самом деле изменение частоты не имеет места, когда речь идет о вторичных волнах, идущих *в направлении распространения света.* Действительно, пусть в газе распространяется плоская монохроматическая волна с частотой ω. Речь идет о частоте в системе отсчета S, в которой газ покоится. Рассмотрим какой-либо движущийся атом. Перейдем в систему отсчета S', в которой атом неподвижен. В системе S' частота распространяющейся плоской волны изменится и будет равна, скажем, ω'. С той же частотой в системе S' возбудятся колебания атома и будут излучаться вторичные сферические волны. При обратном переходе в систему S частота ω' излучаемой сферической волны изменится и будет зависеть от направления излучения. Только для излучения, идущего в направлении первичной волны, получится прежняя частота ω, независимо от того, с какой скоростью и в каком направлении двигался атом.

Таким образом, в направлении распространения первичной волны все атомы будут излучать волны *с одной и той же частотой* ω. С этим и связана возможность регулярного распространения света в газах. Во всех других направлениях движущиеся атомы будут посылать волны с различными частотами. Например, если атом движется в направлении света, то в обратном направлении он будет излучать волны с меньшей частотой. Если же он движется навстречу свету, то частота излучаемой волны в направлении его движения увеличится.

В твердых и жидких телах тепловое движение носит иной характер. В этих случаях атомы движутся *ускоренно*, и рассуждение с переходом к движущейся системе отсчета здесь неприменимо. Атомы совершают колебания около положений равновесия и тем самым *модулируют* поле световой волны. В результате не только сохраняются вторичные волны с прежней частотой, но возникают и волны с новыми частотами. К излучениям с прежними частотами применимо все сказанное выше. С ними связана возможность регулярного распространения световых волн в твердых и жидких средах, а также правильного отражения и преломления их на зеркальных поверхностях тел. Излучения же с изменившимися частотами приводят к появлению в рассеянном свете новых частот.

6. Закончим этот параграф замечанием, которое понадобится нам при выводе формул Френеля с атомистической точки зрения. Если среда однородна и неограниченна, то в ней могут распространяться дипольные колебания в виде бегущей волны

$$\mathbf{p} = \mathbf{p}_0 e^{i(\omega t - \mathbf{kr})}, \tag{68.3}$$

где **p** — дипольный момент атома с радиусом-вектором **r**. Каждый диполь, излучая, теряет энергию. Но эта убыль энергии восполняется за счет энергии, приходящей от других диполей. Излучение других диполей создает в месте нахождения рассматриваемого диполя электрическое поле, которое поддерживает установившиеся гармонические колебания этого диполя. Таким образом, вся среда ведет себя как *замкнутая система*, совершающая *свободные*, а не вынужденные колебания без каких бы то ни было внешних воздействий. Если длина волны велика по сравнению с межатомными расстояниями, то среду можно считать сплошной и характеризовать ее состояние вектором поляризации

$$\mathbf{P} = N\mathbf{p} = \mathbf{P}_0 e^{i(\omega t - \mathbf{kr})}, \tag{68.4}$$

где N — число атомов в единице объема.

§ 69. Вывод формул Френеля в молекулярной оптике

1. Допустим, что однородная изотропная среда граничит с вакуумом вдоль плоскости. Падающая на нее плоская электромагнитная волна возбудит в среде *дипольную волну* (68.3), которую при усреднении по физически бесконечно малым объемам можно рассматривать как *волну поляризации* (68.4). Направления распространения этой и отраженной волн найдутся из условия равенства фазовых скоростей всех волн параллельно границе раздела. Это приводит к равенству *тангенциальных слагающих* волновых векторов $\mathbf{k}_0$, $\mathbf{k}_0'$ и $\mathbf{k}$ падающей, отраженной и прошедшей волн. Кроме того, должны выполняться соотношения $k_0^2 = k_0'^2 = (\omega/c)^2$, так как волна падает на среду из вакуума.

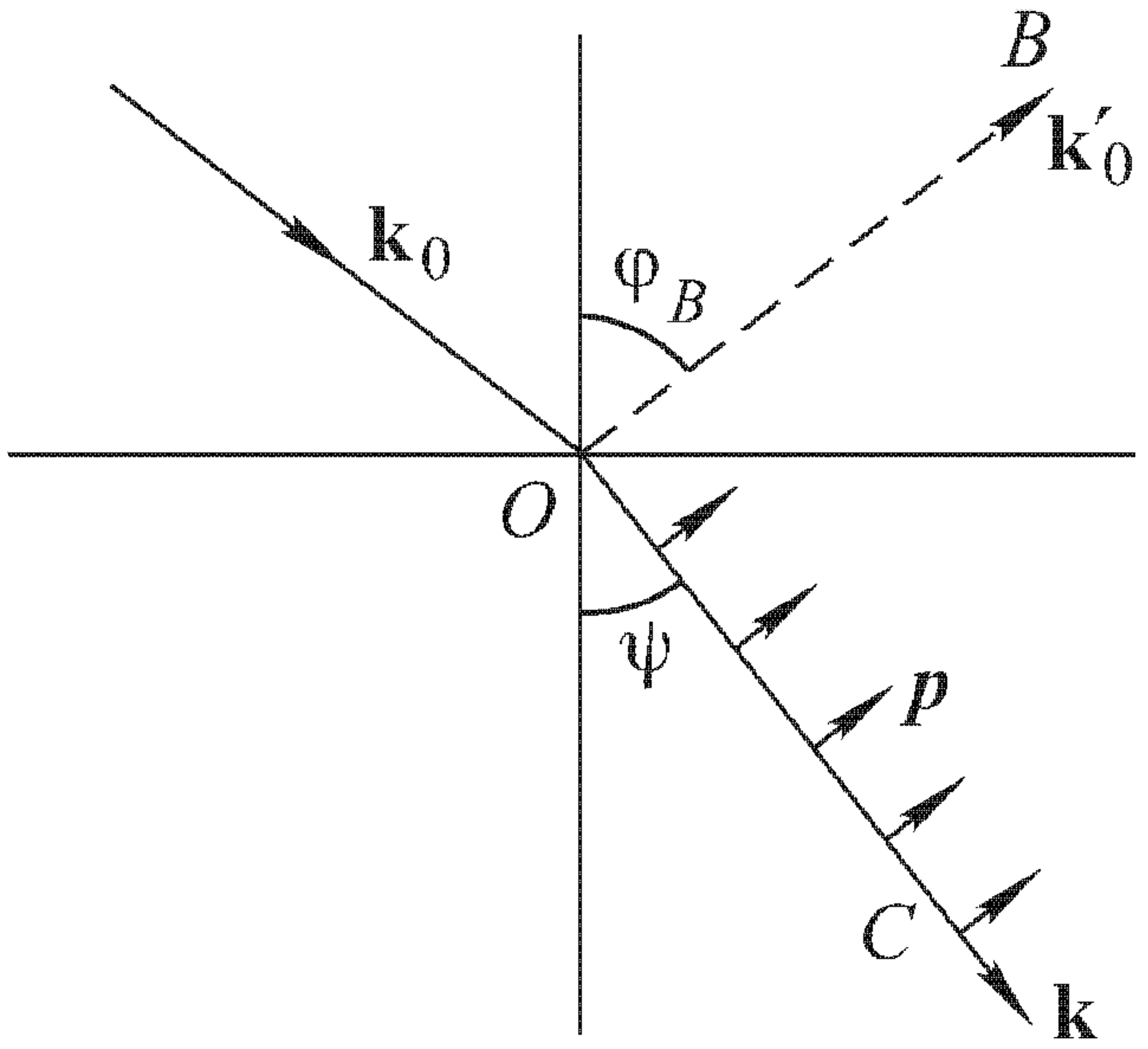


Рис. 251

Пусть теперь электрический вектор падающей волны лежит в плоскости падения (рис. 251). Если волна падает под углом Брюстера, то преломленный луч OC будет перпендикулярен к направлению отраженного луча OB. Ввиду поперечности световых волн электрический вектор в среде перпендикулярен к преломленному лучу OC. Возбуждаемые им дипольные моменты атомов будут также перпендикулярны к OC, а следовательно, параллельны OB. Но вдоль колебаний дипольного момента диполь Герца не излучает. Значит, в направлении OB атомы среды *не излу-*

чают и отраженная волна возникнуть не может. Так просто объясняется закон Брюстера. (См. также задачу 4 к следующему параграфу.)

2. При выводе формул Френеля для отраженного света среду можно считать сплошной, так как межатомные расстояния пренебрежимо малы по сравнению с расстояниями от границы среды до точки наблюдения, где определяется поле отраженной волны. При такой идеализации излучающими центрами будут не отдельные атомы, а элементы объема среды, малые по сравнению с λ^3, но содержащие еще очень много атомов.

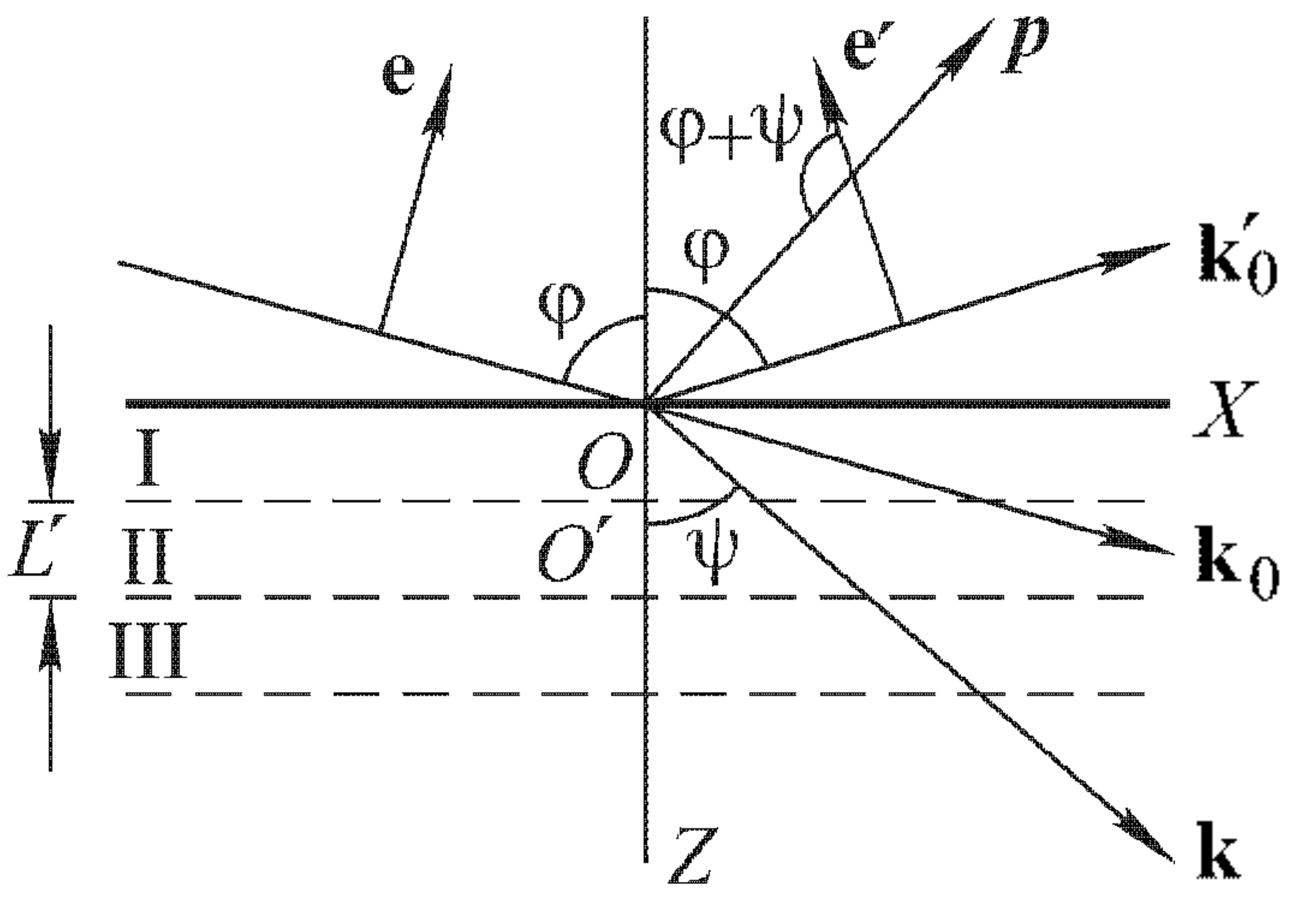


Рис. 252

Выберем координатные оси так же, как это было сделано на рис. 238. Разобьем среду на плоскопараллельные слои I, II, III, ... одинаковой толщины, называемые в дальнейшем *зонами*, по аналогии с кольцевыми зонами Френеля (рис. 252). Каждая плоскопараллельная зона будет излучать в верхнее полупространство плоскую волну с волновым вектором $\mathbf{k}_0'$.

Толщину зоны L' подберем так, чтобы каждая последующая волна запаздывала по фазе на π относительно предыдущей волны. В момент $t = 0$ фаза дипольных колебаний в точке O равна нулю, в точку O' дипольные колебания приходят с фазой $(-k_z L')$. Волна же, излученная в O', приходит в O с дополнительным запаздыванием по фазе на $k_{0z} L'$. Результирующее запаздывание по фазе $(k_z + k_{0z})L'$ по условию должно быть равно π, откуда и находится требуемое значение $L' = \pi/(k_z + k_{0z})$.

Теперь ясно, что электрическое поле $\mathbf{E}^{(r)}$ отраженной волны представится знакопеременным рядом

$$\mathbf{E}^{(r)} = \mathbf{E}_1 - \mathbf{E}_2 + \mathbf{E}_3 - \ldots, \tag{69.1}$$

члены которого представляют поля излучения соответствующих зон. При строгой однородности волны поляризации (68.4) члены ряда (69.1) одинаковы по абсолютной величине. Частичные суммы ряда колеблются между нулем и $\mathbf{E}_1$, а сумма самого ряда не имеет определенного значения.

В действительности реальная волна поляризации *не однородна.* Она имеет *передовой фронт*, перед которым волнового возмущения нет. На самом деле ряд (69.1) содержит конечное число членов, а потому сходится. Для нахождения его суммы мысленно удалим первую зону вместе с излучаемым ею полем и сдвинем оставшуюся среду вверх на расстояние L'. Изменим далее фазы всех диполей среды на одну и ту же величину и притом такую, чтобы диполи, оказавшиеся после смещения на границе раздела, получили те же фазы, какие имели бы в тот же момент времени удаленные с этой границы диполи первой зоны. Ввиду медленности изменения членов ряда (69.1) и тождественности зон, поле излучения среды в верхнем полупространстве в результате этих операций практически останется без изменения. Но теперь оно может быть представлено рядом

$$\mathbf{E}^{(r)} = \mathbf{E}_2 - \mathbf{E}_3 + \mathbf{E}_4 - \ldots \tag{69.2}$$

Складывая его с предыдущим, найдем

$$\mathbf{E}^{(r)} = \frac{1}{2}\,\mathbf{E}_1. \tag{69.3}$$

Таким образом, *напряженность поля излучения среды в верхнем полупространстве равна половине напряженности поля, излучаемого в то же полупространство первой зоной.* Мы пришли к тому же конечному результату, что и в теории кольцевых зон Френеля (см. § 39). Однако здесь наши результаты более убедительны, поскольку соотношение (69.2) для плоских зон лучше обосновано, чем аналогичное соотношение для кольцевых зон Френеля.

3. Прежде чем идти дальше, необходимо сделать следующее замечание. Отраженная волна, излучаемая сплошной средой с непрерывно распределенным вектором поляризации (68.4), в верхнем полупространстве будет всюду плоской и однородной. В случае реальной среды, построенной из дискретных атомов, на плавно меняющееся поле волны вблизи границы среды накладывается «молекулярная рябь», т.е. дополнительное поле, резко меняющееся на расстояниях порядка межатомных расстояний. Это поле может быть представлено в виде суперпозиции волн, быстро затухающих при удалении от границы среды. Пространственный период вдоль оси X таких неоднородных или поверхностных волн порядка расстояния a между атомами. Соответствующая составляющая волнового вектора будет порядка $K_x \approx \approx 2\pi/a$. А так как длина самого вектора $\mathbf{K}$ должна быть равна ω/c, то для его z-составляющей получится $K_z = \sqrt{(\omega/c)^2 - K_x^2} \approx \approx i \cdot 2\pi/a$. Отсюда видно, что при удалении от границы раздела на расстояние a поле неоднородных волн убывает приблизительно в $e^{2\pi} \approx 500$ раз.

Выше указывалось, что замена дискретных излучающих центров (атомов) непрерывно распределенными источниками годится при вычислении поля излучения в верхнем полупространстве на таких расстояниях от границы, которые очень велики по сравнению с a. Приведенная оценка показывает, что это ограничение в действительности должно быть сильно ослаблено. Поле излучения реальной среды не отличается от соответствующего поля ее идеализированной сплошной модели уже на расстояниях от границы порядка a.

4. Дальше при выводе формул Френеля можно было бы пойти двумя путями. Можно было бы рассчитать поле излучения диполей среды в занимаемом ею полупространстве. Это сложно. Мы обойдем вычисления, использовав замечание в конце предыдущего параграфа.

Представим падающую волну как *результат излучения некоторой среды*. Вообразим, что верхнее полупространство на рис. 252 заполнено такой же средой, что и нижнее, а во всем бесконечном пространстве распространяется однородная волна поляризации (68.4) Воображаемая среда оказывает влияние на среду в нижнем полупространстве посредством своего поля излучения. А так как в нижнем полупространстве ничего не изменилось, то поле, излучаемое воображаемой средой в нижнее полупространство, тождественно с полем падающей волны.

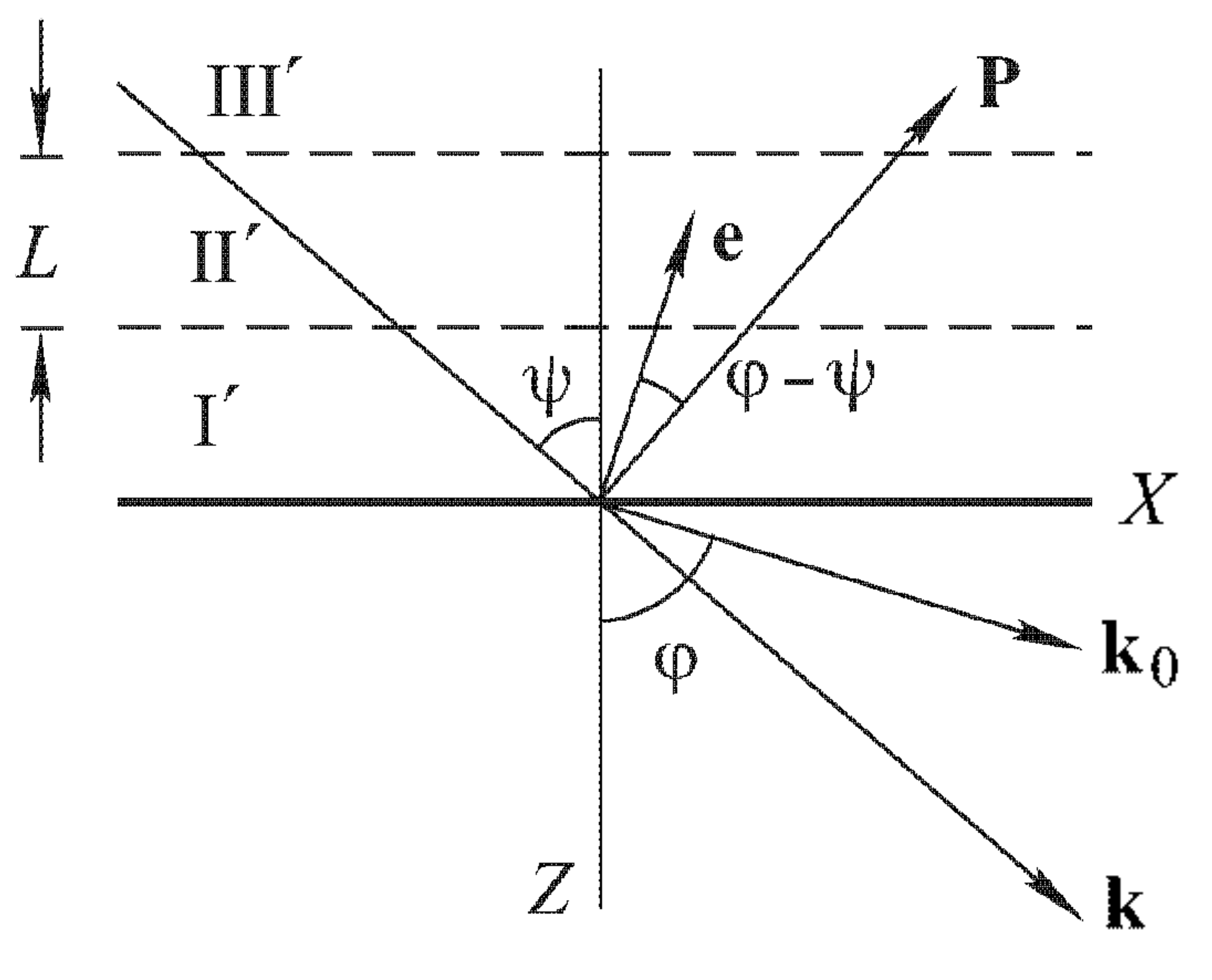


Рис. 253

Разобьем воображаемую среду на плоскопараллельные зоны I′, II′, III′, толщина каждой из которых равна $L = \pi/(k_z - k_{0z})$ (рис. 253). Рассуждая, как и раньше, легко показать, что напряженность поля, излучаемого воображаемой средой в нижнее полупространство, равна половине напряженности поля излучения зоны I'. Вычисление отношения амплитуд отраженной и падающей волн R и $\mathcal{E}$ свелось, таким образом, к сравнению полей излучения зон I и I' на границе раздела $z = 0$.

Если вектор **P** перпендикулярен к плоскости падения, то отношение амплитуд R и $\mathcal{E}$ будет просто равно отношению толщин зон I и I', взятому со знаком минус, поскольку фазы волн, излучаемых этими зонами, на границе раздела противоположны. Это дает первую формулу Френеля:

$$\frac{R_\perp}{\mathcal{E}_\perp} = -\frac{L'}{L} = -\frac{k_z - k_{0z}}{k_z + k_{0z}} = -\frac{\sin(\varphi - \psi)}{\sin(\varphi + \psi)}. \tag{69.4}$$

Если же вектор **P** параллелен плоскости падения, то необходимо еще учесть зависимость поля излучения диполя от угла между дипольным моментом и направлением излучения. Излучение дает только поперечная слагающая дипольного момента. Поэтому величину $-L'/L$ необходимо умножить на отношение проекций вектора **P** на направления единичных векторов $\mathbf{e}'$ и $\mathbf{e}$, перпендикулярных к направлениям распространения отраженной и падающей волн. С учетом поперечности преломленной волны ($\mathbf{kP} = 0$) для этого отношения нетрудно найти $-\cos(\varphi+\psi)/\cos(\varphi-\psi)$ (см. рисунки 252 и 253), а для отношения самих амплитуд $R_{\|}$ и $\mathcal{E}_{\|}$

$$\frac{R_{\|}}{\mathcal{E}_{\|}} = -\frac{R_{\perp}}{\mathcal{E}_{\perp}}\frac{\cos(\varphi+\psi)}{\cos(\varphi-\psi)} = \frac{\operatorname{tg}(\varphi-\psi)}{\operatorname{tg}(\varphi+\psi)}. \tag{69.5}$$

Это — вторая формула Френеля.

Таким же путем может быть разобран случай обратного падения, когда волна падает из среды на ее границу с вакуумом, а также общий случай отражения на границе раздела двух материальных сред.

§ 70. Отступления от формул Френеля

1. Отражение и преломление света на чистой границе раздела двух прозрачных изотропных сред хорошо описывается формулами Френеля. Небольшие отступления от них заметны лишь при отражении под углом Брюстера и вблизи него. С особой отчетливостью они проявляются в существовании следующих двух явлений:

а) при отражении света, поляризованного в плоскости падения, *закон Брюстера соблюдается не строго*; коэффициент отражения не обращается в нуль ни для какого угла падения, хотя при угле Брюстера он и очень мал;

б) свет, линейно поляризованный под углом к плоскости падения, при отражении в малой окрестности угла Брюстера становится поляризованным *эллиптически*, тогда как при строгой справедливости формул Френеля он должен был бы оставаться поляризованным линейно во всех случаях, за исключением случаев полного отражения.

Эти два отступления *не независимы*, а взаимно обусловливают друг друга. Действительно, при точной справедливости формул Френеля

$$\frac{R_{\perp}}{\mathcal{E}_{\perp}} = -\frac{\sin(\varphi-\psi)}{\sin(\varphi+\psi)}, \qquad \frac{R_{\|}}{\mathcal{E}_{\|}} = \frac{\operatorname{tg}(\varphi-\psi)}{\operatorname{tg}(\varphi+\psi)} \tag{70.1}$$

отношение $R_{\perp}/\mathcal{E}_{\perp}$ сохраняло бы знак при всех углах падения, тогда как знак отношения $R_{\|}/\mathcal{E}_{\|}$ менялся бы при переходе через угол Брюстера. Это значит, что во втором случае при переходе через угол Брюстера фаза отражающейся волны скачкообразно менялась бы на π. Никакого нарушения непрерывности $R_{\|}$, как функции угла падения φ, при этом не произошло бы, так как, согласно второй формуле (70.1), эта функция при $\varphi = \varphi_B$ обращается в нуль. Однако функция $R_{||}(\varphi)$ сделалась бы разрывной при $\varphi = \varphi_B$, если бы, как это действительно имеет место, она при этом угле не была равна нулю. Для такого разрыва нет физических оснований. Поэтому следует заключить, что изменение фазы, претерпеваемое $R_{||}(\varphi)$, происходит не скачкообразно при угле $\varphi = \varphi_B$, а в какой-то малой окрестности вокруг угла Брюстера. В пределах этой окрестности фаза меняется непрерывно от 0 до π. В той же окрестности возникает сдвиг фаз между составляющими $\mathbf{E}_{||}$ и $\mathbf{E}_{\perp}$, а потому отраженный свет будет поляризован эллиптически, если только падающий был поляризован линейно. Это показывает, что *второе отступление является следствием первого.*

Обратное также справедливо: *первое отступление вытекает из второго*, если только среды не вращают плоскость поляризации. Это непосредственно очевидно, если заметить, что при соблюдении указанного условия волны $\mathbf{E}_{||}$ и $\mathbf{E}_{\perp}$ сохраняют характер своей поляризации и в соответствии с принципом суперпозиции распространяются независимо друг от друга.

2. Такие отступления от формул Френеля можно объяснить наличием на поверхности отражающей среды тонкого *переходного слоя*, оптические свойства которого отличны от оптических свойств самой среды. В правильности такого объяснения не приходится сомневаться, по крайней мере когда среда свободна от внутренних напряжений. Но на вопрос, как эти слои возникают и каковы их свойства, еще нельзя ответить с полной определенностью.

Простейшим является случай, когда на поверхности отражающей среды существует тонкая пленка постороннего вещества, нанесенная искусственно или возникшая в результате обработки и загрязнения поверхности. Когда толщина пленки велика по сравнению с межатомными расстояниями, ее можно характеризовать *макроскопическими величинами* — толщиной l и показателем преломления n. Пленки с толщиной порядка λ искусственно наносятся на отражающие поверхности для увеличения их поверхностной прозрачности, а также для изготовления интерференционных светофильтров. Если пленка однородна, т.е. величина n одинакова по всей толщине ее, то отражение и преломление света описывается формулами (67.9).

Нас, однако, сейчас интересуют *тонкие пленки*, толщина которых мала по сравнению с λ. В этом случае формулы (67.9) можно упростить, разложив их по степеням l/λ и сохранив только члены первой степени. Для таких пленок вычисление можно обобщить, не вводя предположения об однородности пленки, а предполагая, что при переходе через пленку показатель преломления меняется *непрерывно.* Так поступили Друде (1863–1906) и многие другие авторы. Однако оценка толщин переходных слоев на поверхностях чистых жидкостей и свежих сколов кристаллов показала, что эти толщины *того же порядка*, что и атомные размеры или межатомные расстояния. Применять внутри таких слоев макроскопические уравнения Максвелла и характеризовать их показателем преломления не корректно. Мы получим нужные результаты методами молекулярной оптики, изложенными в предыдущем параграфе.

3. Допустим сначала, что никакого переходного слоя на отражающей поверхности нет. Как выяснено в предыдущем параграфе, для определения поля отраженной волны достаточно рассмотреть излучение первой зоны I (см. рис. 252). Разобьем эту зону на бесконечно тонкие слои толщины $d\zeta$ (рис. 254). Вектор поляризации в точке O' определяется выражением $\mathbf{P}_0 e^{i(\omega t - k_z\zeta)}$. На единицу площади слоя $d\zeta$ приходится дипольный момент $\mathbf{P}_0\, d\zeta\, e^{i(\omega t - k_z\zeta)}$. Колебания из O' приходят в O с запаздыванием по фазе на $k_{0z}\zeta$. При вычислениях слой $d\zeta$ можно мысленно перенести на границу среды, введя соответствующий сдвиг фаз. Дипольный момент, приходящийся на единицу площади так мысленно передвинутого слоя, будет

$$d\mathbf{A}^{(r)} e^{i\omega t} = \mathbf{P}_0 e^{i\omega t} e^{-i(k_{0z}+k_z)\zeta}\, d\zeta.$$

Проинтегрируем это выражение по толщине первой зоны I и результат разделим на два. Тогда найдем, что отраженная волна может быть моделирована полем излучения «эквивалентной плоскости», расположенной на границе среды, на которой совершаются дипольные колебания с поверхностной плотностью амплитуды

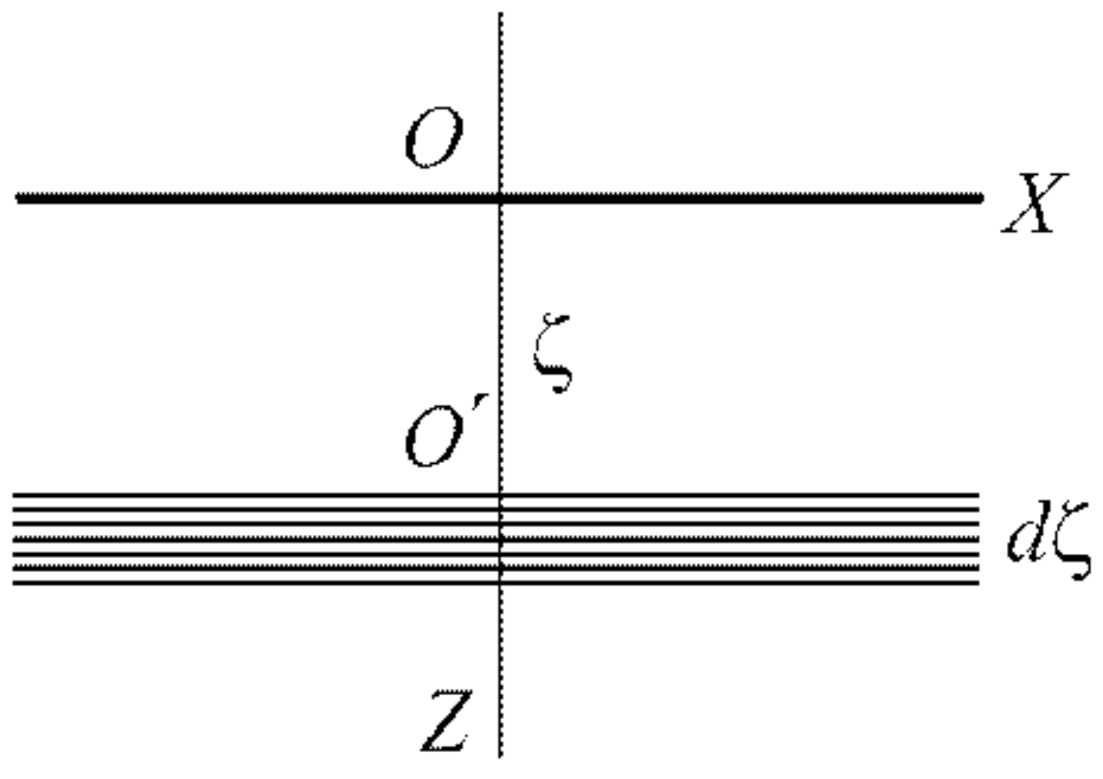


Рис. 254

$$\mathbf{A}_0^{(r)} = \frac{\mathbf{P}_0}{2}\int\limits_0^{L'} e^{-i(k_{0z}+k_z)\zeta}\, d\zeta = i\,\frac{\mathbf{P}_0}{2(k_{0z}+k_z)}\left[e^{-i(k_{0z}+k_z)L'} - 1\right],$$

или, с учетом соотношения $(k_{0z}+k_z)L' = \pi$,

$$\mathbf{A}_0^{(r)} = -i\,\frac{\mathbf{P}_0}{k_{0z}+k_z}. \qquad (70.2\text{а})$$

Рассуждая так же применительно к зоне I′ (см. рис. 253), найдем, что падающая волна может быть моделирована полем излучения другой «эквивалентной плоскости», расположенной также на границе среды. На единицу площади такой плоскости приходится дипольный момент с амплитудой

$$\mathbf{A}_0^{(e)} = -i\,\frac{\mathbf{P}_0}{k_{0z}-k_z}. \qquad (70.2\text{б})$$

4. Учтем теперь наличие переходного слоя. Так как толщина слоя предполагается малой по сравнению с λ, его можно моделировать математической плоскостью. Обозначим через $\boldsymbol{\tau} = \boldsymbol{\tau}_0 e^{i(\omega t - k_{0x}x)}$ дополнительный дипольный момент, вносимый переходным слоем на единицу его площади. Хотя $\boldsymbol{\tau}$ — величина макроскопическая, ею допустимо пользоваться для *сколь угодно тонких переходных слоев* (в том числе и мономолекулярных). Это связано с тем, что введение $\boldsymbol{\tau}$ предполагает усреднение не по малым объемам среды, а *по малым площадкам* отражающей поверхности. Последние же можно выбрать малыми по сравнению с длиной волны, но еще большими по сравнению с межатомными расстояниями.

Величину $\boldsymbol{\tau}_0$ надо добавить к амплитуде $\mathbf{A}_0^{(r)}$, чтобы моделировать поле отраженной волны с учетом излучения переходного слоя. Для моделирования падающей волны ту же величину $\boldsymbol{\tau}_0$ надо добавить к амплитуде $\mathbf{A}_0^{(e)}$, но с противоположным знаком. Действительно, после такого добавления волна, проникающая извне внутрь среды и воздействующая на нее, будет моделирована излучением двух плоскостей, расположенных на границе среды: плоскости с амплитудой дипольных колебаний $(\mathbf{A}_0^{(e)} - \boldsymbol{\tau}_0)$ и плоскости с амплитудой таких же колебаний $\boldsymbol{\tau}_0$. Амплитуды $\boldsymbol{\tau}_0$ и $-\boldsymbol{\tau}_0$ компенсируются. Остается *только одна амплитуда* $\mathbf{A}_0^{(e)}$, как если бы переходный слой не оказывал никакого влияния на среду вдали от ее границы. Но это как раз и необходимо, чтобы поляризация среды вне переходного слоя представлялась однородной волной (68.4).

Таким образом, вместо (70.2а) и (70.2б) получаются выражения

$$\mathbf{A}^{(r)} = -i\,\frac{\mathbf{P}_0}{k_z + k_{0z}} + \boldsymbol{\tau}_0, \qquad \mathbf{A}^{(e)} = i\,\frac{\mathbf{P}_0}{k_z - k_{0z}} - \boldsymbol{\tau}_0, \tag{70.3}$$

которыми и надо пользоваться при вычислении поля падающей и отраженной волн. При этом, разумеется, надо учесть, что для излучения существенны не сами амплитуды (70.3), а только их составляющие, *перпендикулярные к направлению излучения*. Коэффициенты Френеля $R_\perp/\mathcal{E}_\perp$ и $R_\parallel/\mathcal{E}_\parallel$ найдутся как отношения таких составляющих. Ограничиваясь линейными членами по k_0, в результате простых, но несколько длинных вычислений получим

$$\frac{R_\perp}{\mathcal{E}_\perp} = -\frac{\sin(\varphi - \psi)}{\sin(\varphi + \psi)}(1 + 2ik_0\gamma_x\cos\varphi), \tag{70.4}$$

$$\frac{R_\parallel}{\mathcal{E}_\parallel} = \frac{\operatorname{tg}(\varphi - \psi)}{\operatorname{tg}(\varphi + \psi)}\left(1 + 2ik_0\cos\varphi\,\frac{\gamma_x\cos^2\psi - \gamma_z\sin^2\varphi}{\cos^2\psi - \sin^2\varphi}\right). \tag{70.5}$$

Здесь введены обозначения:

$$\gamma_x = \frac{\tau_{0x}}{P_{0x}} \equiv \frac{\tau_{0y}}{P_{0y}}, \qquad \gamma_z = \frac{\tau_{0z}}{P_{0z}}. \tag{70.6}$$

Параметры γ_x и γ_z имеют размерность длины. В изложенной теории ими и характеризуются свойства переходного слоя. Теория справедлива при выполнении условий

$$\gamma_x k_0 \ll 1, \quad \gamma_z k_0 \ll 1, \quad \text{или} \quad \gamma_x/\lambda \ll 1, \quad \gamma_z/\lambda \ll 1.$$

Квадратичные члены $(\gamma_x/\lambda)^2$ и $(\gamma_z/\lambda)^2$ в ней отброшены. В этом приближении все коэффициенты при малых величинах γ_x/λ и γ_z/λ не зависят от λ. Это значит, что вычисление параметров γ_x и γ_z есть *чисто электростатическая задача*. Для фактического вычисления самих параметров в молекулярной теории необходимо знать молекулярную структуру среды и переходного слоя.

5. Состояние поляризации отраженного света полностью характеризуется комплексным отношением $R_{\parallel}/R_{\perp}$. Если $\mathcal{E}_{\perp} = \mathcal{E}_{\parallel}$, то

$$\frac{R_{\parallel}}{R_{\perp}} = -\frac{\cos(\varphi+\psi)}{\cos(\varphi-\psi)} + 2ik_0(\gamma_z - \gamma_x)\frac{\cos\varphi\sin^2\varphi}{\cos^2(\varphi-\psi)}. \tag{70.7}$$

При угле Брюстера это отношение *чисто мнимое*, т.е. отраженный свет поляризован *эллиптически*. Одна из главных осей эллипса колебаний перпендикулярна к плоскости падения, а другая лежит в этой плоскости. Отношение этих осей называется *коэффициентом эллиптичности света*, отраженного под углом Брюстера. Он может быть и положительным, и отрицательным, так как разность фаз δ между составляющими $R_{\parallel}$ и $R_{\perp}$ при угле Брюстера принято считать равной $+\pi/2$. При таком соглашении $R_{\parallel}/R_{\perp} = \rho e^{i\pi/2} = i\rho$. Для ρ из формулы (70.7) нетрудно получить

$$\rho = \frac{\pi}{\lambda}\sqrt{n^2+1}\,(\gamma_z - \gamma_x), \tag{70.8}$$

где λ означает длину волны в вакууме. Положительному коэффициенту эллиптичности соответствует левая эллиптическая поляризация отраженного света, отрицательному — правая (предполагается, что падающий свет линейно поляризован под углом 45° к плоскости падения, т.е. $\mathcal{E}_{\parallel} = \mathcal{E}_{\perp}$). Эллиптическая поляризация называется *правой*, если для глаза, смотрящего против распространения света, движение по эллипсу совершается по часовой стрелке. В противоположном случае она называется *левой*.

Для всех исследованных жидкостей с чистой отражающей поверхностью коэффициент ρ положителен. Для обыкновенной воды он отрицателен. Однако, как показал Рэлей, при очищении поверхности воды величина ρ меняет знак, обращаясь в нуль при некоторой промежуточной степени очищения. Для предельно чистой поверхности воды Рэлей нашел $\rho = +0{,}00042$. В табл. 6 приведены значения ρ для различных жидкостей (при $\lambda = 546$ нм).

Таблица 6

Вещество	$\rho \cdot 10^5$	n	$l_{\text{мин}}/\lambda$
Вода	42	1,33	0,00057
Метиловый спирт	46	1,33	0,00062
Этиловый спирт	92	1,33	0,00124
Четыреххлористый углерод	107	1,46	0,00103
Бензол	135	1,51	0,00117
Толуол	144	1,50	0,00101
Ортоксилол	137	1,51	0,00119

В последнем столбце табл. 6 приведены значения минимальной толщины $l_{\text{мин}}$ переходного слоя, которые по теории Друде достаточны для получения наблюдаемой эллиптичности отраженного света (см. задачу 3 к этому параграфу). Хотя эта теория и макроскопическая, приведенная оценка дает правильный порядок величины $l_{\text{мин}}$. Из нее видно, что переходные слои порядка одного или нескольких межатомных расстояний уже достаточны, чтобы вызвать наблюдаемые отступления от формул Френеля. Это с несомненностью указывает на то, что, наряду с переходными слоями, вызванными загрязнением, обработкой, абсорбцией газов и пр., существуют и переходные слои, обусловленные *молекулярной структурой* самой отражающей среды вблизи ее поверхности.

Трудно указать причины возникновения таких переходных слоев. Вот некоторые из возможных причин. 1. Электрическое поле световой волны, действующее на приграничные молекулы среды, может отличаться от действующего поля для остальных молекул. 2. Анизотропные молекулы, если среда построена из таковых, могут ориентироваться вблизи поверхности в некоторых преимущественных направлениях. 3. Плотность среды в тонком приграничном слое может отличаться от плотности остальной среды.

ЗАДАЧИ

1. Считая переходный слой сплошным, вычислить для него параметры γ_x и γ_z.

Решение. Задача сводится к нахождению вектора поляризации $\mathbf{P}(z)$ среды и переходного слоя в однородном электрическом поле $\mathbf{E}$. Среду будем характеризовать диэлектрическими проницаемостями $\varepsilon_x(z)$ и $\varepsilon_z(z)$ в направлениях параллельно и перпендикулярно к слою. Координатные оси можно выбрать так, чтобы $E_y = D_y = 0$. В силу симметрии векторы $\mathbf{E}$ и $\mathbf{D}$ не могут зависеть от x и y. Поэтому уравнения $\operatorname{rot}\mathbf{E} = 0$ и $\operatorname{div}\mathbf{D} = 0$ переходят в

$$\frac{dE_x}{dz} = \frac{dD_z}{dz} = 0.$$

Отсюда следует, что во всем пространстве E_x и D_z постоянны, т.е. не зависят от z. Вектор поляризации в переходном слое будет

$$P_x = \frac{\varepsilon_x - 1}{4\pi}E_z, \quad P_z = \frac{\varepsilon_z - 1}{4\pi}E_z = \frac{1}{4\pi}\left(1 - \frac{1}{\varepsilon_z}\right)D_z.$$

Чтобы выделить дополнительную поляризацию $\Delta\mathbf{P}$ переходного слоя, вычтем отсюда вектор поляризации самой однородной среды:

$$\Delta P_x = \frac{\varepsilon_x - \varepsilon}{4\pi}E_x, \quad \Delta P_z = \frac{1}{r\pi}\left(\frac{1}{\varepsilon} - \frac{1}{\varepsilon_z}\right)D_z,$$

где ε — диэлектрическая проницаемость этой среды. Для нахождения τ_{0x} и τ_{oz} полученные выражения надо проинтегрировать по толщине l переходного слоя. Выполнив это и введя обозначения

$$p = \int\limits_0^l \varepsilon_x(z)\,dz, \quad q = \int\limits_0^l \frac{dz}{\varepsilon_z(z)},$$

найдем

$$\gamma_x = \frac{p - ln^2}{n^2 - 1}, \quad \gamma_z = \frac{l - qn^2}{n^2 - 1}. \tag{70.9}$$

Если подставить эти значения в (70.4) и (70.5), то получатся *формулы Друде.*

2. Считая переходный слой макроскопическим и однородным, получить выражение для коэффициента эллиптичности при угле Брюстера.

Ответ.

$$\rho = \frac{\pi}{\lambda}\sqrt{n^2 + 1}\frac{(n^2 - n'^2)(n'^2 - 1)}{(n^2 - 1)n'^2}\,l, \tag{70.10}$$

где n' — показатель преломления переходного слоя.

3. Считая эллиптичность ρ положительной, найти значение n', чтобы при заданных n и l величина ρ получилась максимальной. Найти минимальное значение для l, чтобы при заданном n получилось заданное значение ρ.

Ответ.

$$n' = \sqrt{n}, \quad l_{\text{мин}} = \frac{\rho}{\pi\sqrt{n^2+1}}\frac{n+1}{n-1}\lambda. \qquad (70.11)$$

4. Изотропная среда построена из анизотропных молекул. Ее макроскопическая изотропия обусловлена беспорядочными изменениями направлений молекул в результате теплового движения. Дипольный момент анизотропной молекулы, возбуждаемый полем световой волны, вообще говоря, не совпадает по направлению с электрическим полем. Пусть электрический вектор падающей волны лежит в плоскости падения, а сама волна падает под углом Брюстера. Электрический вектор в среде хотя и перпендикулярен к преломленному лучу, но дипольные моменты молекул, вообще говоря, имеют составляющие вдоль луча. Такие молекулы, очевидно, будут излучать в направлении отраженного луча. Будет ли в этом случае выполняться закон Брюстера? (См. § 69, п. 1.)

Решение. Закон Брюстера останется справедливым. Для макроскопически изотропной среды составляющие дипольных моментов молекул, параллельные преломленному лучу, меняются хаотически во времени. В силу этого излучения отдельных молекул в направлении отраженного луча некогерентны. Они не могут привести к правильному отражению, а только к диффузному рассеянию света. Но если в тонком поверхностном слое анизотропные молекулы преимущественно ориентированы в некотором направлении, то излучения таких молекул становятся частично когерентными, и появляется отраженный свет (см. § 69).

ГЛАВА VI

ОПТИКА МЕТАЛЛОВ

§ 71. Уравнения Максвелла и волны в металлах

1. Как и в диэлектриках, распространение электромагнитных волн в металлах описывается уравнениями Максвелла:

$$\begin{gathered}\operatorname{rot}\mathbf{H}=\frac{1}{c}\,\dot{\mathbf{D}}+\frac{4\pi}{c}\mathbf{j},\\ \operatorname{rot}\mathbf{E}=-\frac{1}{c}\,\dot{\mathbf{B}},\\ \operatorname{div}\mathbf{D}=4\pi\rho,\\ \operatorname{div}\mathbf{B}=0.\end{gathered}\tag{71.1}$$

По сравнению с диэлектриками добавляется член с током проводимости $\mathbf{j}$. В статических и низкочастотных полях ток обусловлен движением практически одних только *свободных электронов*. Движение связанных электронов, а тем более атомных ядер в этой области спектра не играет никакой роли. Но уже в инфракрасной области, где лежат собственные частоты колебаний атомных ядер, движения последних начинают существенно влиять на оптическое поведение металлов (резонанс). В дальнейшем, в видимой и ультрафиолетовой областях спектра, в колебания вовлекаются и *связанные электроны*.

Смещения связанных зарядов вызывают *поляризацию металлов*, аналогичную поляризации диэлектриков. Появляется *ток поляризации*. Однако в высокочастотных полях нет существенной разницы между движениями свободных и связанных частиц. Нерационально связывать ток проводимости с движением только свободных, а ток поляризации — только связанных электронов. Этим понятиям надо дать точные определения, в соответствии с тем, как они используются в теории.

Можно поступить следующим образом. Вынужденное колебание заряженной частицы (электрона или атомного ядра) в поле световой волны можно разложить на колебание, происходящее в фазе (или противофазе) с электрическим полем, и колебание, сдвинутое относительно него по фазе на 90°. Ток, обусловленный первыми колебаниями, условимся называть *током проводимости* и обозначать его плотность через $\mathbf{j}$. Токи проводимости вызывают поглощение электромагнитных волн. Токи же, обусловленные колебаниями, сдвинутыми по фазе на 90°, назовем *токами поляризации.* С ними не связано поглощение волн. Плотность тока поляризации можно представить в виде $\mathbf{j}_{\text{пол}} = \dot{\mathbf{P}}$, где $\mathbf{P}$ — вектор поляризации. На последнее соотношение можно смотреть как на определение вектора $\mathbf{P}$. Тем самым в металлах раскрывается и точный смысл вектора индукции $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$, который входит в уравнения (71.1).

Магнитные свойства вещества, не играющие существенной роли в оптической области спектра, мы учитывать не будем. Металлы будем считать *оптически изотропными*, хотя все металлы (за исключением, конечно, жидких) и имеют кристаллическую структуру. Однако кристаллы кубической системы оптически изотропны. Другие металлы, как правило, *макроскопически изотропны*, так как они состоят из множества хаотически ориентированных кристалликов, размеры которых малы по сравнению с длиной волны. Для изотропных металлов материальные уравнения имеют вид

$$\mathbf{D} = \varepsilon'\mathbf{E}, \tag{71.2}$$

$$\mathbf{j} = \sigma\mathbf{E}. \tag{71.3}$$

Величина σ называется *электрической проводимостью*, а ε' — *диэлектрической проницаемостью* металла. (Мы пользуемся обозначением ε', сохраняя ε для обозначения комплексной величины, вводимой ниже.) Обе величины ε' и σ являются функциями частоты ω. Поэтому в уравнениях (71.2) и (71.3) поле $\mathbf{E}$ должно предполагаться монохроматическим. Немонохроматические поля надо разлагать на монохроматические составляющие и применять принцип суперпозиции.

С наличием проводимости σ связано *поглощение света* в металлах. С точки зрения излагаемой здесь формальной теории поглощение света есть не что иное, как превращение электромагнитной энергии в *джоулево тепло.*

Материальные уравнения (71.2) и (71.3) дают лишь грубое описание оптических свойств металлов (см. § 74). В ряде вопросов, в особенности для коротких волн (ультрафиолетовые, видимые и короткие инфракрасные лучи), они приводят к выводам, не совсем согласующимся с опытом. Более удовлетворительная теория должна основываться на *квантовой теории металлов.*

Однако изложение такой теории далеко выходит за рамки этой книги.

2. Для монохроматических полей $\dot{\mathbf{D}} = i\omega\mathbf{D} = i\omega\varepsilon'\mathbf{E}$, так что правая часть первого уравнения (71.1) преобразуется в $\frac{1}{c}(i\omega\varepsilon' + 4\pi\sigma)\mathbf{E}$. Если ввести величину

$$\varepsilon = \varepsilon' - i\,\frac{4\pi\sigma}{\omega} \equiv \varepsilon' - i\varepsilon'', \tag{71.4}$$

называемую *комплексной диэлектрической проницаемостью*, то система уравнений (71.1) примет вид

$$\operatorname{rot}\mathbf{H} = \frac{i\omega}{c}\,\varepsilon\mathbf{E}, \quad \operatorname{rot}\mathbf{E} = -\frac{i\omega}{c}\,\mathbf{H}, \tag{71.5}$$

т.е. формально совпадет с соответствующими уравнениями для диэлектриков. Граничные условия также имеют одинаковый вид: они требуют непрерывности тангенциальных компонент векторов **E** и **H** на границе раздела сред. Как следствие таких граничных условий и уравнений (71.5), получается непрерывность нормальных компонент вектора $\varepsilon\mathbf{E}$ (см. § 63). Поэтому *любое соотношение оптики прозрачных сред, полученное из уравнений* (71.5) *и граничных условий с помощью линейных вещественных операций, может быть формально перенесено в оптику металлов и других поглощающих сред простой заменой вещественной величины* ε *на комплексную.* Требуется только дополнительное исследование физического содержания и смысла полученного соотношения.

Для характеристики оптических свойств металлов применяется также *комплексный показатель преломления* ν. Он определяется соотношением

$$\nu^2 = \varepsilon = \varepsilon' - i\varepsilon''. \tag{71.6}$$

Полагая

$$\nu = n - i\varkappa, \tag{71.7}$$

где n и $\varkappa$ вещественны и существенно положительны, получим

$$n^2 - \varkappa^2 = \varepsilon', \quad 2n\varkappa = \varepsilon''. \tag{71.8}$$

Величина n называется *главным показателем преломления* металла. Величину $\varkappa$ называют *главным показателем затухания.* Этот термин надо предпочесть обычно употребляемому термину «показатель поглощения», так как *затухание волны может происходить и без поглощения.* Примером может служить плазма, когда частота ω меньше так называемой *плазменной частоты* (см. § 87). Для нее величина ε вещественна, но *отрицательна*, т.е. $\varepsilon'' = 0$. В этом случае поглощения нет, но есть затухание, так как $\sqrt{\varepsilon}$ — величина чисто мнимая, а потому $\varkappa \neq 0$.

Если среда однородна, то в ней могут распространяться плоские монохроматические волны вида (5.3). Для них должны выполняться соотношения

$$\mathbf{E} = -\frac{c}{\omega\varepsilon}\,[\mathbf{kB}], \quad \mathbf{B} = \frac{c}{\omega}\,[\mathbf{kE}], \tag{71.9}$$

$$\mathbf{k}^2 = \frac{\omega^2}{c^2}\,\varepsilon = \frac{\omega^2}{c^2}\,(\varepsilon' - i\varepsilon''). \tag{71.10}$$

В поглощающей среде волновой вектор $\mathbf{k}$ всегда комплексный, а соответствующая плоская волна *всегда неоднородна.* Это естественно, так как при наличии поглощения плоская волна не может распространяться без затухания. Положим

$$\mathbf{k} = \mathbf{k}' - i\mathbf{k}'', \tag{71.11}$$

где $\mathbf{k}'$ и $\mathbf{k}''$ — вещественные векторы. Тогда на основании (71.10)

$$k'^2 - k''^2 = \frac{\omega^2}{c^2}\varepsilon', \quad 2(\mathbf{k}'\mathbf{k}'') = \frac{\omega^2}{c^2}\varepsilon''. \tag{71.12}$$

Вектор $\mathbf{k}'$ указывает направление распространения *плоскостей равных фаз.* В направлении вектора $\mathbf{k}''$ убывает амплитуда волны. В общем случае *плоскости равных фаз и плоскости равных амплитуд не перпендикулярны между собой.* Перпендикулярность всегда имеет место только для непоглощающих сред, когда $\varepsilon'' = 0$.

§ 72. Геометрические законы отражения и преломления света на границе металла

1. Пусть из вакуума на плоскую границу металла падает плоская монохроматическая однородная волна, распространяющаяся вдоль волнового вектора $\mathbf{k}_1$ (рис. 255). При этом возникнет однородная отраженная волна с волновым вектором $\mathbf{k}_1'$ и неоднородная волна, прошедшая в металл. Комплексный волновой вектор прошедшей волны обозначим через $\mathbf{k}$ (без индекса). Как было показано ранее, из граничных условий получаются соотношения

$$k_{1x} = k_{1x}' = k_x. \tag{72.1}$$

Из них следует, что геометрические законы отражения света от металлов такие же, что и для непоглощающих сред. Различие есть лишь в законах преломления.

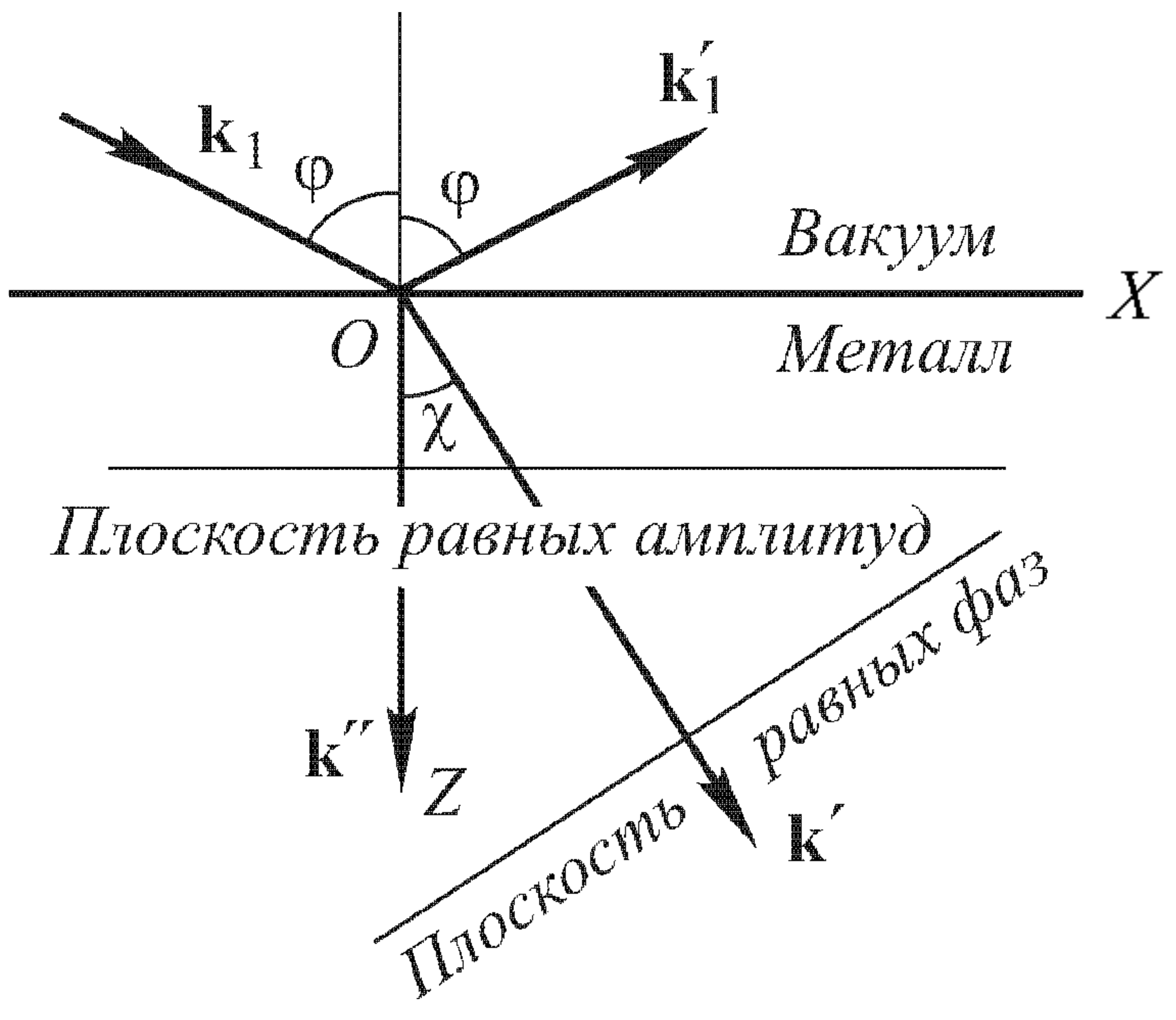


Рис. 255

Прежде всего отметим, что плоскости равных амплитуд прошедшей волны параллельны границе металла. Действительно, представим комплексный вектор $\mathbf{k}$ в виде (71.11). Из (72.1) следует, что тангенциальная составляющая вектора $\mathbf{k}$ вещественна, а потому вектор $\mathbf{k}''$ перпендикулярен к поверхности металла. Это и доказывает наше утверждение. Прошедшая волна затухает в направлении вектора $\mathbf{k}''$. Поэтому вектор $\mathbf{k}''$ надо направить вниз, т.е. в сторону металла, так как затухание волны в металле должно идти в этом, а не в противоположном направлении. В сторону металла должен быть направлен и вектор $\mathbf{k}'$, поскольку угол между векторами $\mathbf{k}'$ и $\mathbf{k}''$ острый, как это следует из второго соотношения (71.12).

Вектор $\mathbf{k}'$ перпендикулярен к плоскости равных фаз прошедшей волны. Угол χ, образуемый этим вектором с положительным направлением оси Z, называется *вещественным углом преломления*. Отношение

$$\frac{\sin\varphi}{\sin\chi} = n_\varphi, \tag{72.2}$$

вообще говоря, зависит от угла падения φ. Оно положительно и называется *показателем преломления*. Так как $k_x = k'_x = k'\sin\chi$, то из соотношения (72.1) следует

$$k' = k_1\frac{\sin\varphi}{\sin\chi} = \frac{\omega}{c}n_\varphi. \tag{72.3}$$

По аналогии с (72.3) введем другую положительную величину $\varkappa_\varphi$, определяемую соотношением

$$k'' = \frac{\omega}{c}\varkappa_\varphi. \tag{72.4}$$

Ее называют *показателем затухания*. Физический смысл показателя затухания легко установить, рассмотрев выражение для поля прошедшей волны. Таким путем не представляет труда установить, что на глубине

$$z = h_\varphi = \frac{\lambda_0}{4\pi\varkappa_\varphi}, \tag{72.5}$$

где λ_0 — длина волны в вакууме, интенсивность света (пропорциональная квадрату амплитуды) убывает в e раз. Величина h_φ называется *глубиной проникновения света в металл*. Таким образом, *показатель затухания $\varkappa_\varphi$ можно определить как отношение длины световой волны в вакууме к умноженной на 4π глубине проникновения света в металл.*

Для металлов показатель затухания в видимой области спектра обычно порядка единицы. Например, для золота при нормальном падении $\varkappa_\varphi = 2{,}82$, $h_\varphi = \lambda_0/(4\pi\varkappa_\varphi) = \lambda_0/35{,}4$. Отсюда находим, что на протяжении длины волны интенсивность света в золоте убывает в $\exp(\lambda_0/h_\varphi) \approx 2{,}4\cdot 10^{15}$ раз.

Пленки металлов с толщиной порядка длины волны, как правило, практически непрозрачны для света. Об оптических свойствах металлов обычно судят *по отраженному свету*. Тем не менее необходимо изучить законы проникновения света в металл, так как без этого нельзя понять и законы отражения. Свет, отраженный от металла (как и от диэлектрика), возникает в результате интерференции когерентных вторичных волн, излучаемых электронами и атомными ядрами металла. Но вторичные волны, очевидно, возбуждаются падающей волной, проникшей в металл. Если бы поле в металл совсем не проникало, то отражение света было бы невозможно.

2. Выразим теперь n_φ и $\varkappa_\varphi$ через оптические константы металла n, $\varkappa$ и угол падения φ. Подставляя выражения (72.3) и (72.4) в формулы (71.12), получим

$$n_\varphi^2 - \varkappa_\varphi^2 = \varepsilon', \quad 2n_\varphi\varkappa_\varphi\cos\chi = \varepsilon''. \tag{72.6}$$

Сравнивая эти выражения с (71.8), находим

$$n_\varphi^2 - \varkappa_\varphi^2 = n^2 - \varkappa^2, \quad n_\varphi\varkappa_\varphi\cos\chi = n\varkappa. \tag{72.7}$$

Кеттелер (1839–1900), впервые получивший эти соотношения, назвал их *главными уравнениями распространения световых волн в металлах и поглощающих средах*. Они показывают, что величины

$$a \equiv n_\varphi^2 - \varkappa_\varphi^2, \quad b \equiv 2n_\varphi\varkappa_\varphi\cos\chi \tag{72.8}$$

не зависят от угла падения φ, а только от рода металла, его физического состояния и от длины световой волны. Величины a и b называются *инвариантами Кеттелера* и могут служить для характеристики оптических свойств металлов (вместо n и $\varkappa$).

Если $\varphi = 0$, то $\chi = 0$, и уравнения (72.7) дают $n_\varphi = \pm n$, $\varkappa_\varphi = \pm\varkappa$. Здесь надо взять знак плюс, так как по определению величины n_φ и $\varkappa_\varphi$ существенно положительны. Таким образом, при нормальном падении показатели преломления и затухания n_φ и $\varkappa_\varphi$ принимают свои главные значения n и $\varkappa$.

Для нахождения n_φ и $\varkappa_\varphi$ как функций угла падения φ перепишем второе уравнение (72.7) в виде

$$n_\varphi^2\varkappa_\varphi^2\cos^2\chi = n_\varphi^2\varkappa_\varphi^2(1 - \sin^2\chi) = n_\varphi^2\varkappa_\varphi^2 - \varkappa_\varphi^2\sin^2\varphi = \frac{1}{4}\,b^2.$$

Решая его совместно с первым уравнением, найдем

$$\begin{aligned} n_\varphi^2 &= \frac{1}{2}\left[\pm\sqrt{(a - \sin^2\varphi)^2 + b^2} + (a + \sin^2\varphi)\right], \\ \varkappa_\varphi^2 &= \frac{1}{2}\left[\pm\sqrt{(a - \sin^2\varphi)^2 + b^2} - (a - \sin^2\varphi)\right]. \end{aligned} \tag{72.9}$$

Знаки перед квадратными корнями должны быть одинаковы в обеих формулах, чтобы разность $n_\varphi^2 - \varkappa_\varphi^2$ была равна a. Кроме того, n_φ и $\varkappa_\varphi$ должны быть непрерывными функциями угла φ. Для действительно поглощающих сред инвариант b не равен нулю, как это видно из выражения $b = 2n\varkappa = \varepsilon''$. Поэтому квадратный корень в формулах (72.9) не может обращаться в нуль и менять знак при изменении угла φ. Но при $\varphi = 0$ величины n_φ^2 и $\varkappa_\varphi^2$ могут быть существенно положительными тогда и только тогда, когда оба квадратных корня в (72.9) взяты со знаком плюс. Значит, знак плюс следует брать и при любых значениях угла φ. Таким образом, для дейсвительно поглощающих сред

$$\begin{aligned} n_\varphi^2 &= \frac{1}{2}\left[\sqrt{(a-\sin^2\varphi)^2+b^2}+(a+\sin^2\varphi)\right], \\ \varkappa_\varphi^2 &= \frac{1}{2}\left[\sqrt{(a-\sin^2\varphi)^2+b^2}-(a-\sin^2\varphi)\right]. \end{aligned} \tag{72.10}$$

В частности, при $\varphi = 0$

$$n^2 = \frac{1}{2}(\sqrt{a^2+b^2}+a), \quad \varkappa^2 = \frac{1}{2}(\sqrt{a^2+b^2}-a). \tag{72.11}$$

Особые случаи в выборе знака могут иметь место только при $b \equiv 2n\varkappa = 0$, т.е. либо при $n = 0$, $\varkappa \neq 0$, либо при $n \neq 0$, $\varkappa = 0$. В обоих случаях диэлектрическая проницаемость вещественна, т.е. среда непоглощающая. Однако величины n_φ и $\varkappa_\varphi$ сохраняют смысл и для таких случаев. Например, теорию полного отражения на границе прозрачных сред (см. § 66) можно представить как частный случай теории, изложенной в этом параграфе.

§ 73. Формулы Френеля. Измерение оптических констант металлов

1. Формулы Френеля (65.7) или (65.8) применимы и для металлов, если под $\sin\psi$ и $\cos\psi$ понимать комплексные величины

$$\sin\psi = \frac{\sin\varphi}{\nu}, \quad \cos\psi = \frac{1}{\nu}\sqrt{\nu^2-\sin^2\varphi}. \tag{73.1}$$

Здесь надо взять то значение квадратного корня, которое имеет отрицательную мнимую часть — только тогда неоднородная волна, проникшая в поглощающую среду, будет затухать при удалении от границы раздела.

При нормальном падении

$$\frac{R_\perp}{\mathcal{E}_\perp} = -\frac{R_\parallel}{\mathcal{E}_\parallel} = -\frac{\nu-1}{\nu+1} = -\frac{(n-1)-i\varkappa}{(n+1)-i\varkappa}. \tag{73.2}$$

Для отражательной способности металла получаем

$$R = \left|\frac{R_\perp}{\mathcal{E}_\perp}\right|^2 = \frac{(n-1)^2+\varkappa^2}{(n+1)^2+\varkappa^2}. \tag{73.3}$$

2. При отражении от металла оба отношения $R_\perp/\mathcal{E}_\perp$ и $R_\parallel/\mathcal{E}_\parallel$, вообще говоря, комплексны, т.е. появляются *скачки фаз*. Они, как правило, *различны* для составляющих $\mathbf{E}_\parallel$ и $\mathbf{E}_\perp$. *Если падающий свет поляризован линейно под углом к плоскости падения, то отраженный свет будет поляризован эллиптически.* Исследуя эллиптическую поляризацию отраженного света, можно определить оптические константы металла n и $\varkappa$. На этом основан *метод Друде*, излагаемый ниже.

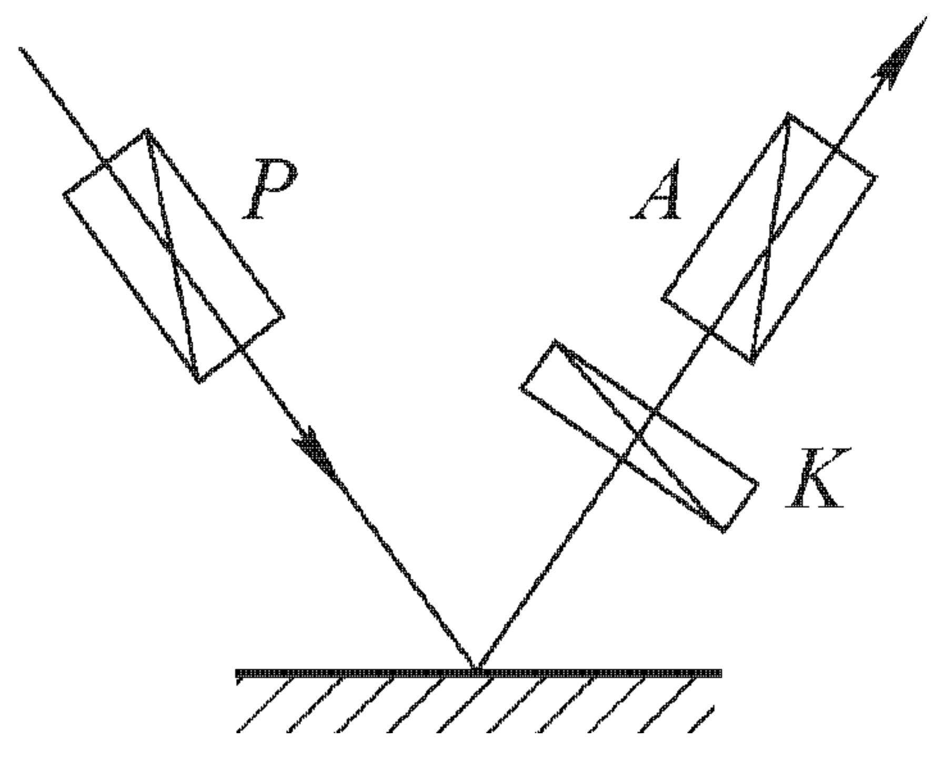


Рис. 256

Световой луч, пройдя через поляризатор P (рис. 256), поляризуется линейно. Для простоты расчета предположим, что азимут поляризации равен 45° ($\mathcal{E}_\perp = \mathcal{E}_\parallel$). Обобщение на случай произвольного азимута не встречает затруднений. Отраженный луч сначала проходит через компенсатор K, а затем через анализатор A. Изменяя установку компенсатора и вращая анализатор вокруг направления отраженного луча, можно погасить отраженный луч. В этом случае после прохождения через компенсатор свет становится поляризованным линейно. Азимут его поляризации называется *азимутом восстановленной линейной поляризации* отраженного света. Компенсатором K можно измерить разность фаз Δ между $\mathbf{E}_\parallel$ и $\mathbf{E}_\perp$ отраженной волны, а анализатором A — азимут β ее восстановленной линейной поляризации. По этим данным можно вычислить оптические константы металла n и $\varkappa$.

Действительно, из формул Френеля (65.8) при $\mathcal{E}_\perp = \mathcal{E}_\parallel$ получаем

$$\frac{R_\parallel}{R_\perp} = -\frac{\cos(\varphi + \psi)}{\cos(\varphi - \psi)}.$$

Очевидно,

$$\frac{R_\parallel}{R_\perp} = re^{i\Delta}, \qquad r = \operatorname{tg}\beta.$$

Отсюда

$$\frac{1 - re^{i\Delta}}{1 + re^{i\Delta}} = \frac{\cos\varphi\cos\psi}{\sin\varphi\sin\psi} = \frac{\sqrt{\nu^2 - \sin^2\varphi}}{\operatorname{tg}\varphi\sin\varphi}.$$

Далее,

$$\nu^2 - \sin^2\varphi = (n - i\varkappa)^2 - \sin^2\varphi = n^2 - \varkappa^2 - 2in\varkappa - \sin^2\varphi.$$

На основании (72.7) и (72.2)

$$\nu^2 - \sin^2\varphi = n_\varphi^2 - \varkappa_\varphi^2 - 2in_\varphi\varkappa_\varphi\cos\chi - n_\varphi^2\sin^2\chi = (n_\varphi\cos\chi - i\varkappa_\varphi)^2.$$

на его поверхности. В табл. 7 приведены для ориентировки значения $\varkappa$ и n некоторых металлов, полученные по методу Друде (λ = 589,3 нм). Отражательная способность R вычислена по формуле (73.3).

3. Запишем формулу (73.3) в виде

$$R = \frac{(n^2+\varkappa^2+1)-2n}{(n^2+\varkappa^2+1)+2n}. \tag{73.8}$$

Как видно из табл. 7, почти для всех металлов $2n$ мало по сравнению с $n^2+\varkappa^2+1$. Благодаря этому отражательная способность металлов *очень велика* — для многих металлов она близка к единице. «Металлический блеск», присущий металлам, объясняется их большой отражательной способностью. Так как $\varkappa$ и n изменяются в зависимости от длины волны, некоторые металлы — в особенности медь и золото — обладают резко выраженным цветом. Металл кажется нам, например, красным, если он сильнее всего отражает красные лучи. Поэтому, грубо говоря, поверхностная окраска металла является дополнительной к цвету лучей, проходящих через тонкие металлические пленки. Например, очень тонкие золотые пленки в проходящем свете кажутся зелеными.

Таблица 7

Металл	$\varkappa$	n	R, %
Натрий	2,61	0,05	97,5
Серебро, цельное	3,64	0,18	95,1
Магний, цельный	4,42	0,37	93,1
Золото, цельное	2,82	0,37	84,9
Золото, электролитическое	2,83	0,47	81,5
Ртуть	4,41	1,62	75,4
Медь, цельная	2,62	0,64	73,2
Никель, цельный	3,32	1,79	62,0
Никель, электролитический	3,48	2,01	62,0
Никель, гальванически распыленный	1,97	1,30	43,3
Железо, гальванически распыленное	1,63	1,51	32,6

Для слабо поглощающих тел отражательная способность определяется почти исключительно главным показателем преломления n и практически не зависит от $\varkappa$. К таким телам относятся красящие вещества. В этом случае поверхностная окраска тел обусловлена не избирательным отражением, а *избирательным поглощением* в слое красящего вещества. Поэтому окраска тел кажется одинаковой в отраженном и проходящем свете. Слои лаковых красок свободны от неоднородностей, которые сделали бы эти слои мутными. Сквозь слои таких красок видны детали поверхности тела. Свет от источника попадает на поверхность тела и на ней рассеивается. Таким образом, на пути к глазу он дважды проходит через поглощающий слой краски, претерпевая при этом избирательное поглощение. Свет, отраженный от передней поверхности лака, может дать слабый беловатый оттенок. Если поверхность слоя лака гладкая, то такой беловатый оттенок практически ограничивается областью зеркального отражения. Клеевые краски искусственно делаются мутными путем введения в них различных неоднородностей. Свет, падающий на слой краски, рассеивается на этих неоднородностях, не достигая поверхности тела. Таким образом, значительная доля света рассеивается верхними слоями краски, не претерпевая избирательного поглощения. Благодаря этому появляется заметный беловатый оттенок.

Различие в происхождении поверхностной окраски слабо и сильно поглощающих сред можно иллюстрировать и следующим примером. Покроем поверхность прозрачной стеклянной пластинки слоем фиолетовых чернил. Чернила, пока они не засохли, являются «слабо поглощающей средой». Они кажутся фиолетовыми как в отраженном, так и в проходящем свете: окраска обусловлена избирательным поглощением света, проходящего через слой чернил и испытывающего рассеяние в нем. Когда чернила засохнут, они превращаются в «сильно поглощающую среду». В проходящем свете они по-прежнему кажутся фиолетовыми — окраска обусловлена избирательным поглощением. В отраженном же свете засохший слой чернил приобретает дополнительный желтовато-зеленый металлический блеск, вызванный избирательным отражением.

Зависимость оптических констант многих металлов от длины волны выражена весьма резко. Так, серебро, характеризующееся в видимой области большим коэффициентом отражения (около 95 %), имеет в ультрафиолете резко выраженную область плохого отражения и большой прозрачности — вблизи $\lambda = 316$ нм отражательная способность серебра снижается до 4,2 %, т.е. становится такой же, как у стекла. Вуд показал, что тонкие пленки щелочных металлов прозрачны в ультрафиолетовой части спектра, но совершенно не пропускают видимых лучей. Ему удалось даже обнаружить угол Брюстера при отражении ультрафиолетовых лучей от этих металлов.

Для многих поглощающих тел отражательная способность имеет резко выраженный максимум в некоторой, иногда весьма узкой, области спектра. Путем многократных отражений от таких тел можно получить лучи с довольно высокой степенью монохроматичности. Такие лучи называются *остаточными*. Этот метод применяли Рубенс (1865–1922) и его сотрудники для получения монохроматических инфракрасных лучей.

4. В заключение остановимся на отражательной способности металлов для длинных волн (радиоволны, инфракрасные лучи). В этой области проводимость σ практически не зависит от частоты ω и равна своему статическому значению σ_0. Как показывает формула (71.4), мнимая часть ε'' диэлектрической проницаемости с уменьшением частоты растет. Для длинных волн вещественной частью ε' можно пренебречь, полагая

$$\varepsilon(\omega) = -i\frac{4\pi\sigma_0}{\omega}. \tag{73.9}$$

Тогда из соотношений (71.8) получаем

$$n = \varkappa = \sqrt{\frac{\varepsilon''}{2}} = \sqrt{\frac{2\pi\sigma_0}{\omega}} = \sqrt{\sigma_0 T} = \sqrt{\frac{\lambda\sigma_0}{c}}, \tag{73.10}$$

где T — период колебаний, λ — длина волны в вакууме.

Соотношения (73.10) впервые были получены Друде. Для видимого света они не оправдываются. Так, для меди $\sigma_0 = 5{,}14 \cdot 10^{17}$ с$^{-1}$. Полагая $\lambda = 589{,}3$ нм $= 5{,}893 \cdot 10^{-5}$ см, из формул (73.10) найдем $n = \varkappa \approx 33$, тогда как опыт дает $n = 0{,}64$, $\varkappa = 2{,}62$. Соотношениями (73.10) можно пользоваться только для длинных волн. В этом случае (73.8) переходит в

$$R = \frac{2n^2 - 2n + 1}{2n^2 + 2n + 1} \approx 1 - \frac{2}{n}\left(1 - \frac{1}{n} + \ldots\right),$$

или

$$R = 1 - \frac{2}{\sqrt{\sigma_0 T}}\left(1 - \frac{1}{\sqrt{\sigma_0 T}} + \ldots\right). \tag{73.11}$$

Эта формула проверялась Гагеном и Рубенсом для инфракрасных лучей ($\lambda = 4$; 8; 12; 25,5 мкм). Они нашли, что при $\lambda > 5$ мкм формула (73.11) начинает оправдываться количественно. Для радиоволн величина $1/\sqrt{\sigma_0 T}$ очень близка к нулю, а отражательная способность — к единице: металлы отражают радиоволны практически полностью.

ЗАДАЧА

В каких случаях возможно полное отражение света от среды?

Решение. Если свет поляризован перпендикулярно к плоскости падения, то условие полного отражения будет

$$\left|\frac{R_\perp}{\mathcal{E}_\perp}\right|^2 = \frac{\cos\varphi - \nu\cos\psi}{\cos\varphi + \nu\cos\psi} \cdot \frac{\cos\varphi - \nu^*\cos^*\psi}{\cos\varphi + \nu^*\cos^*\psi} = 1,$$

или $\cos\varphi(\nu\cos\varphi + \nu^*\cos^*\psi) = 0$. Пользуясь формулой (73.4), ему можно придать вид

$$n_\varphi \cos\varphi \cos\chi = 0. \tag{73.12}$$

Такое же условие получится и в случае света, поляризованного в плоскости падения. Полное отражение может наступить в одном из трех случаев:

1) $\cos\varphi = 0$; 2) $n_\varphi = 0$; 3) $\cos\chi = 0$. Первый случай соответствует скользящему падению. Во втором случае $n_\varphi = 0$, а потому, ввиду (72.7), $n\varkappa = 0$. В третьем случае $\cos\chi = 0$ и, следовательно, $n\varkappa$ также равно нулю. Итак, если $\varphi \neq \pi/2$, то для полного отражения необходимо, чтобы $n\varkappa = 0$. Это условие может осуществиться в двух случаях: 1) $n = 0$, $\varkappa \neq 0$; 2) $n \neq 0$, $\varkappa = 0$.

Случай 1. $n = 0$, $\varkappa \neq 0$. Отражение будет полным при любом угле падения. Действительно, в рассматриваемом случае $\nu = n - i\varkappa = -i\varkappa$, и формулы Френеля принимают вид

$$\frac{R_\perp}{\mathcal{E}_\perp} = \frac{\cos\varphi + i\varkappa\cos\psi}{\cos\varphi - i\varkappa\cos\psi}, \quad \frac{R_\parallel}{\mathcal{E}_\parallel} = -\frac{i\varkappa\cos\varphi + \cos\psi}{-i\varkappa\cos\varphi + \cos\psi}.$$

В них

$$\cos\psi = \sqrt{1 - \sin^2\psi} = \sqrt{1 - \frac{\sin^2\varphi}{\nu^2}} = \sqrt{1 + \left(\frac{\sin\varphi}{\varkappa}\right)^2},$$

откуда видно, что $\cos\psi$ — величина всегда вещественная. В знаменателях формул Френеля стоят величины, комплексно-сопряженные с числителями. Следовательно, $|R_\perp/\mathcal{E}_\perp|^2 = |R_\parallel/\mathcal{E}_\parallel|^2 = 1$. В рассматриваемом случае $\varepsilon = \nu^2 = -\varkappa^2 < 0$, т.е. среда непоглощающая с отрицательной диэлектрической проницаемостью. Такой случай осуществляется при отражении радиоволн от плазмы (см. § 87).

Случай 2. $n \neq 0$, $\varkappa = 0$. В этом случае $\nu = n$, $\varepsilon = n^2 > 0$, т.е. среда прозрачна. Полное отражение будет иметь место при $n < 1$, если φ превосходит предельный угол $\varphi_0 = \arcsin n$ (см. § 66).

§ 74. Аномальный скин-эффект и эффективная диэлектрическая проницаемость

1. Проникновение электромагнитной волны в тонкий поверхностный слой металла есть частный случай *скин-эффекта*, рассмотренного нами в т. III, § 144. Самый слой, в который проникает электромагнитное поле, называется *скин-слоем*. Толщина скин-слоя определяется формулой (72.5). Она выводится на основе макроскопических уравнений Максвелла (71.5). Из тех же уравнений следует, что напряженность поля в скин-слое убывает *экспоненциально*. Такой скин-слой называется *нормальным*.

Для применимости макроскопических уравнений Максвелла необходимо, чтобы *межатомные расстояния были малы по сравнению не только с длиной волны, но и с толщиной скин-слоя*. Это условие можно считать выполненным для всех металлов. Более жестким является условие применимости понятия диэлектрической проницаемости $\varepsilon(\omega)$, как оно было введено в § 71. Там было учтено, что электроны и ионы, движением которых создаются токи проводимости и поляризации, движутся в электрическом поле, которое меняется во времени, но не было принято во внимание его изменение в пространстве. Это допустимо, когда *средняя длина свободного пробега электронов мала по сравнению с расстояниями, на которых заметно меняется напряженность электрического поля, т.е. по сравнению с длиной волны и толщиной скин-слоя.* Только тогда электрон от столкновения до столкновения движется практически в однородном поле. Если же средняя длина свободного пробега электрона порядка или больше толщины скин-слоя или длины волны, то результаты § 71 требуют пересмотра. Понятие диэлектрической проницаемости $\varepsilon(\omega)$ может потерять смысл. Тогда напряженность поля и ток будут убывать в глубь металла *не экспоненциально*, а по более сложному закону. Соответствующий скин-эффект называется *аномальным.*

Если воспользоваться значениями $\varkappa$ из табл. 7, то легко убедиться, что у всех металлов, приведенных в этой таблице, величина $h = \lambda/(4\pi\varkappa)$ для видимого света порядка 10^{-6} см. Того же порядка при комнатных температурах и средняя длина свободного пробега электронов. Это указывает на аномальный характер скин-эффекта. Только для плохих проводников, у которых длина свободного пробега меньше, скин-эффект при комнатных температурах нормальный. В области же низких температур, где средняя длина свободного пробега достигает для очень чистых образцов значений порядка $10^{-2} - 10^{-3}$ см, об использовании теории нормального скин-эффекта, основанной на понятии диэлектрической проницаемости, не может быть и речи. Поскольку характер скин-эффекта определяется соотношением между средней длиной свободного пробега электрона и толщиной скин-слоя, полная теория аномального скин-эффекта должна строиться на основе *кинетики электронов*, в ее квантовой форме.

2. Однако независимо от того, является ли скин-эффект нормальным или аномальным, отражение света возникает в результате излучения электромагнитных волн токами, текущими в поверхностном слое металла и возбуждаемыми падающей волной. Механизм отражения света от металлов вполне аналогичен соответствующему механизму для диэлектриков, разобранному в § 68 и 69.

В случае нормального скин-эффекта плотность полного тока убывает в глубь металла по экспоненциальному закону. В случае аномального скин-эффекта это не так. Однако если толщина скин-слоя много меньше длины волны, конкретный закон изменения плотности тока в поверхностном слое может лишь слабо сказаться на отражении света, так как в этом случае фазы источников вторичных волн, распределенных в поверхностном слое, практически одинаковы по всей его толщине. Поэтому при вычислении поля отраженной волны действительное распределение полного тока в скин-слое может быть заменено распределением, в котором плотность тока убывает экспоненциально. Такая замена эквивалентна введению вместо ε *эффективной диэлектрической проницаемости металла* $\varepsilon_{\text{эф}}$.

Пользуясь величиной $\varepsilon_{\text{эф}}$, можно вычислять поле отраженной волны так, как если бы скин-эффект был нормальным, а металл имел диэлектрическую проницаемость $\varepsilon = \varepsilon_{\text{эф}}$. Однако если скин-эффект аномальный, пользоваться $\varepsilon_{\text{эф}}$ для вычисления поля внутри металла нельзя: термин «эффективная» понимается в смысле: *эффективная в отношении отражения.* Эффективная диэлектрическая проницаемость, очевидно, может быть введена и в случае более толстых скин-слоев. Однако в этих случаях она, вообще говоря, зависит от состояния поляризации падающей волны и от угла падения. Для теоретического вычисления $\varepsilon_{\text{эф}}$ надо было бы решить задачу об отражении света методами кинетики и статистики, что далеко выходит за рамки этой книги.

ГЛАВА VII

КРИСТАЛЛООПТИКА

§ 75. Плоские волны в кристаллах

1. Большинство кристаллов *оптически анизотропны*, т.е. их оптические свойства в разных направлениях не одинаковы. Наиболее важным проявлением этой анизотропии является *двойное лучепреломление в кристаллах.* Изучение этого явления и связанной с ним поляризации света составит основное содержание настоящей главы.

Фундаментальные уравнения Максвелла справедливы без всяких изменений и в кристаллических средах. В отсутствие электрических зарядов и токов они имеют вид

$$\operatorname{rot}\mathbf{H} = \frac{1}{c}\,\dot{\mathbf{D}}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\,\dot{\mathbf{H}}. \tag{75.1}$$

Но материальные уравнения усложняются. Изотропные среды характеризуются скалярной диэлектрической проницаемостью $\varepsilon(\omega)$. Для характеристики оптических свойств кристаллов требуются девять величин $\varepsilon_{ik}(\omega)$, образующих *тензор диэлектрической проницаемости*, или *диэлектрический тензор.* Он вводится посредством соотношений

$$D_j = \sum_k \varepsilon_{jk} E_k \qquad (j, k = x, y, z). \tag{75.2}$$

Для прозрачных кристаллов, как можно показать, исходя из закона сохранения энергии, диэлектрический тензор *симметричен*, т.е. $\varepsilon_{ij} = \varepsilon_{ji}$ (см. § 80). Разумеется, в различных системах координат компоненты диэлектрического тензора имеют разные значения. При переходе от одной системы координат к другой они преобразуются как компоненты всякого тензора. Благодаря тензорной связи между **D** и **E** направления этих векторов в кристаллах, вообще говоря, не совпадают.

2. Если среды прозрачны и однородны, то в них могут распространяться плоские монохроматические волны. Каждую из них можно записать в виде

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{kr})}, \quad \mathbf{H} = \mathbf{H}_0 e^{i(\omega t - \mathbf{kr})}, \quad \mathbf{D} = \mathbf{D}_0 e^{i(\omega t - \mathbf{kr})}. \tag{75.3}$$

Рассмотрим сначала свойства таких волн, которые вытекают из одних только фундаментальных уравнений (75.1) без использования материальных уравнений (75.2). Как и в случае изотропных сред,

$$\operatorname{rot} \mathbf{H} = -i\,[\mathbf{kH}], \quad \dot{\mathbf{D}} = -i\omega \mathbf{D}, \ldots \tag{75.4}$$

Подставив эти выражения в (75.1) и вводя единичный вектор волновой нормали $\mathbf{N}$ по формуле $\mathbf{k} = \frac{\omega}{v}\mathbf{N}$, получим

$$\mathbf{D} = -\frac{c}{v}\,[\mathbf{NH}], \quad \mathbf{H} = \frac{c}{v}\,[\mathbf{NE}], \tag{75.5}$$

где v — нормальная скорость волны, т.е. фазовая скорость, с которой распространяется волновой фронт в направлении волновой нормали $\mathbf{N}$. Присоединим к этим формулам еще выражение для вектора Пойнтинга

$$\mathbf{S} = \frac{c}{4\pi}\,[\mathbf{EH}]. \tag{75.6}$$

Фундаментальное значение этого вектора в кристаллооптике состоит в том, что он определяет *направление световых лучей*, т.е. линий, вдоль которых происходит распространение энергии света. Чтобы убедиться в этом, достаточно показать, что направление $\mathbf{S}$ совпадает с направлением вектора групповой скорости $\mathbf{u} = d\omega/d\mathbf{k}$ (см. § 8). Это будет сделано в § 81. *В кристаллах векторы* $\mathbf{S}$ *и* $\mathbf{N}$, *вообще говоря, не совпадают по направлению.* Именно с этим связано двойное лучепреломление, а также коническая рефракция.

Из формул (75.5) видно, что *векторы* $\mathbf{D}$ *и* $\mathbf{H}$ *взаимно перпендикулярны*. Кроме того, *они перпендикулярны к волновой нормали* $\mathbf{N}$, т.е. параллельны фронту волны. Значит, *плоские волны в кристалле поперечны в отношении векторов* $\mathbf{D}$ *и* $\mathbf{H}$. Однако *в общем случае они не поперечны в отношении вектора* $\mathbf{E}$. *Четыре вектора* $\mathbf{E}$, $\mathbf{D}$, $\mathbf{N}$, $\mathbf{S}$ *лежат в одной плоскости, перпендикулярной к вектору* $\mathbf{H}$. Взаимное расположение этих векторов показано на рис. 257. Из него видно, что *заданием в кристалле направления вектора* $\mathbf{E}$ (*или* $\mathbf{D}$) *однозначно определяется направление вектора* $\mathbf{D}$ (*или* $\mathbf{E}$) *и с точностью до* 180° — *направления всех остальных векторов* $\mathbf{H}$, $\mathbf{N}$, $\mathbf{S}$. *Опре-*

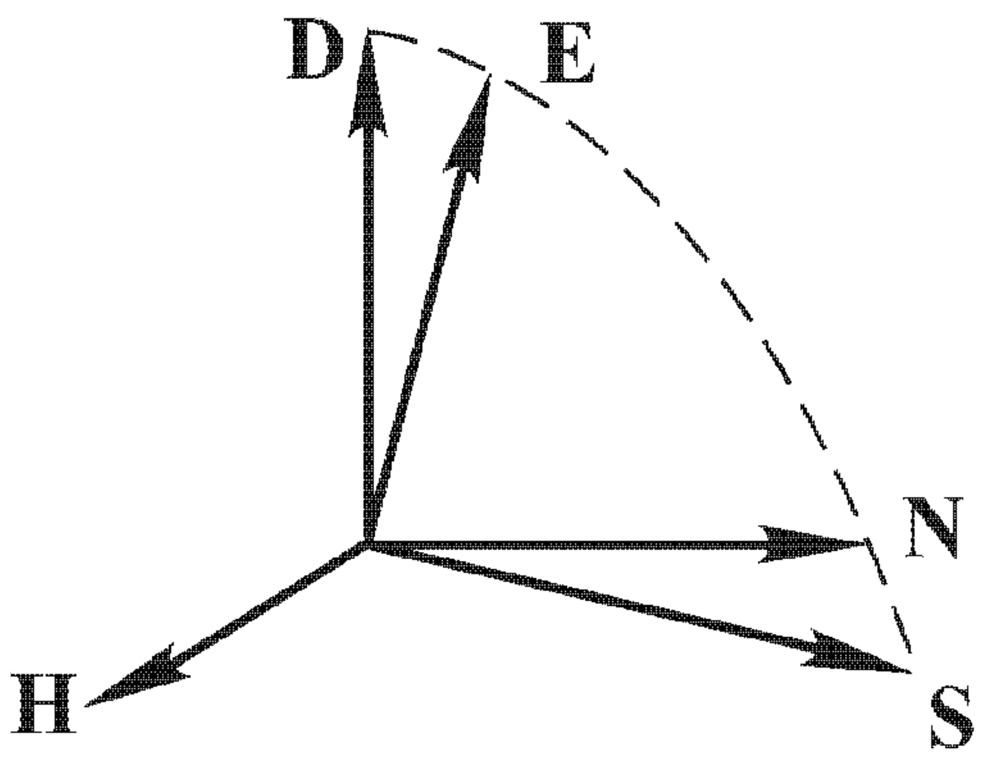


Рис. 257

деляется также величина нормальной скорости волны v. Действительно, если задано направление вектора $\mathbf{E}$, то уравнение (75.2) однозначно определит направление вектора $\mathbf{D}$, а с ним и плоскость $(\mathbf{E}, \mathbf{D})$, к которой перпендикулярен вектор $\mathbf{H}$. Тем самым с точностью до 180° определится направление $\mathbf{H}$, а следовательно, и направления векторов $\mathbf{N}$ и $\mathbf{S}$. Исключение составляет случай, когда направления векторов $\mathbf{E}$ и $\mathbf{D}$ совпадают. Тогда всякая прямая, перпендикулярная к вектору $\mathbf{E}$, может служить направлением магнитного поля. Те же рассуждения применимы и к случаю, когда задано направление вектора $\mathbf{D}$. Для определения v исключим из (75.5) вектор $\mathbf{H}$:

$$\mathbf{D} = \frac{c^2}{v^2}\mathbf{E} - \frac{c^2}{v^2}(\mathbf{N}\mathbf{E})\mathbf{N}. \tag{75.7}$$

Так как $(\mathbf{D}\mathbf{N}) = 0$, то отсюда скалярным умножением на $\mathbf{D}$ находим

$$v^2 = c^2\frac{(\mathbf{D}\mathbf{E})}{D^2}. \tag{75.8}$$

Таким образом, электрический вектор ($\mathbf{E}$ или $\mathbf{D}$) в кристалле в известном смысле является *главным.* Это и понятно, так как именно он определяет электрическую поляризацию среды, а возбуждение последней составляет сущность процесса распространения электромагнитных волн в материальных средах.

§ 76. Оптически одноосные кристаллы

1. Простейшими оптическими свойствами обладают *оптически одноосные кристаллы,* которые к тому же имеют наибольшее практическое значение. Поэтому имеет смысл особо выделить этот простейший частный случай. Оптически одноосными называются кристаллы, свойства которых обладают симметрией вращения относительно некоторого направления, называемого *оптической осью кристалла.*

Разложим электрические векторы $\mathbf{E}$ и $\mathbf{D}$ на составляющие $\mathbf{E}_{\parallel}$, $\mathbf{D}_{\parallel}$ вдоль оптической оси и составляющие $\mathbf{E}_{\perp}$, $\mathbf{D}_{\perp}$, перпендикулярные к ней. Тогда

$$\mathbf{D}_{\parallel} = \varepsilon_{\parallel}\mathbf{E}_{\parallel}, \quad \mathbf{D}_{\perp} = \varepsilon_{\perp}\mathbf{E}_{\perp}, \tag{76.1}$$

где $\varepsilon_{\parallel}$ и $\varepsilon_{\perp}$ — постоянные, называемые *продольной* и *поперечной диэлектрическими проницаемостями кристалла.* К этим двум величинам и сводится диэлектрический тензор одноосного кристалла. К оптически одноосным кристаллам относятся все кристаллы тетрагональной, гексагональной и ромбоэдрической систем. Кристаллы кубической системы являются вырожденным случаем их. Для них $\varepsilon_{\parallel} = \varepsilon_{\perp}$. Поэтому *кристаллы кубической* *системы в оптическом отношении ведут себя как оптически изотропные тела.* Плоскость, в которой лежат оптическая ось кристалла и нормаль $\mathbf{N}$ к фронту волны, называется *главным сечением кристалла.* Главное сечение — это не какая-то определенная плоскость, а целое семейство параллельных плоскостей.

Рассмотрим теперь два частных случая.

Случай 1. Вектор $\mathbf{D}$ перпендикулярен к главному сечению кристалла. В этом случае $\mathbf{D} \equiv \mathbf{D}_\perp$, а потому $\mathbf{D} = \varepsilon_\perp \mathbf{E}$. Кристалл ведет себя как изотропная среда с диэлектрической проницаемостью $\varepsilon_\perp$. Для нее

$$\mathbf{D} = \varepsilon_\perp \mathbf{E}, \tag{76.2}$$

и из уравнений (75.5) получаем

$$D = \frac{c}{v}H, \quad H = \frac{c}{v}E, \tag{76.3}$$

или

$$\varepsilon_\perp E = \frac{c}{v}H, \quad H = \frac{c}{v}E, \tag{76.4}$$

откуда

$$v = v_\perp \equiv v_o = \frac{c}{\sqrt{\varepsilon_\perp}}. \tag{76.5}$$

Таким образом, *если электрический вектор перпендикулярен к главному сечению, то скорость волны не зависит от направления ее распространения.* Такая волна называется *обыкновенной.*

Случай 2. Вектор $\mathbf{D}$ лежит в главном сечении. Так как вектор $\mathbf{E}$ лежит также в главном сечении (см. рис. 257), то $\mathbf{E} = \mathbf{E}_N + \mathbf{E}_D$, где $\mathbf{E}_N$ — составляющая этого вектора вдоль $\mathbf{N}$, а $\mathbf{E}_D$ — вдоль $\mathbf{D}$. Из векторного произведения $[\mathbf{NE}]$ составляющая $\mathbf{E}_N$ выпадает. Поэтому вторую формулу (75.5) можно записать в виде

$$\mathbf{H} = \frac{c}{v}\,[\mathbf{NE}_D].$$

Очевидно,

$$E_D = \frac{\mathbf{ED}}{D} = \frac{E_\parallel D_\parallel + E_\perp D_\perp}{D} = \frac{(D_\parallel^2/\varepsilon_\parallel) + (D_\perp^2/\varepsilon_\perp)}{D},$$

или

$$E_D = D\left(\frac{\sin^2\alpha}{\varepsilon_\parallel} + \frac{\cos^2\alpha}{\varepsilon_\perp}\right) = D\left(\frac{N_\perp^2}{\varepsilon_\parallel} + \frac{N_\parallel^2}{\varepsilon_\perp}\right),$$

где α — угол между оптической осью и волновой нормалью (рис. 258). Если ввести обозначение

$$\frac{1}{\varepsilon} = \frac{N_\perp^2}{\varepsilon_\parallel} + \frac{N_\parallel^2}{\varepsilon_\perp}, \tag{76.6}$$

то получится $D = \varepsilon E_D$, и мы придем к соотношениям

$$\varepsilon E_D = \frac{c}{v}H, \quad H = \frac{c}{v}E_D, \tag{76.7}$$

формально тождественным с соотношениями (76.4). Роль величины $\varepsilon_\perp$ теперь играет величина ε, определяемая выражением (76.6). Поэтому нормальная скорость волны будет определяться выражением

$$v = \frac{c}{\sqrt{\varepsilon}} = c\sqrt{\frac{N_\perp^2}{\varepsilon_\|} + \frac{N_\|^2}{\varepsilon_\perp}} \equiv v_\|. \tag{76.8}$$

Она меняется с изменением направления волновой нормали **N**. По этой причине волну, электрический вектор которой лежит в главном сечении кристалла, называют *необыкновенной. Зависимость ее скорости $v_\|$ от направления распространения обусловлена тем, что с изменением направления волновой нормали* **N** *меняется угол между электрическим вектором и оптической осью кристалла.* Когда $N_\perp = 0$, т.е. необыкновенная волна распространяется вдоль оптической оси кристалла, то из формулы (76.8) получаем $v = c/\sqrt{\varepsilon_\perp} = v_\perp \equiv v_o$. В этом случае, как это очевидно заранее, нет разницы между обыкновенной и необыкновенной волнами. Если же $N_\| = 0$, т.е. необыкновенная волна распространяется перпендикулярно к оптической оси, то скорость волны будет равна

$$v = v_e \equiv c/\sqrt{\varepsilon_\|}. \tag{76.9}$$

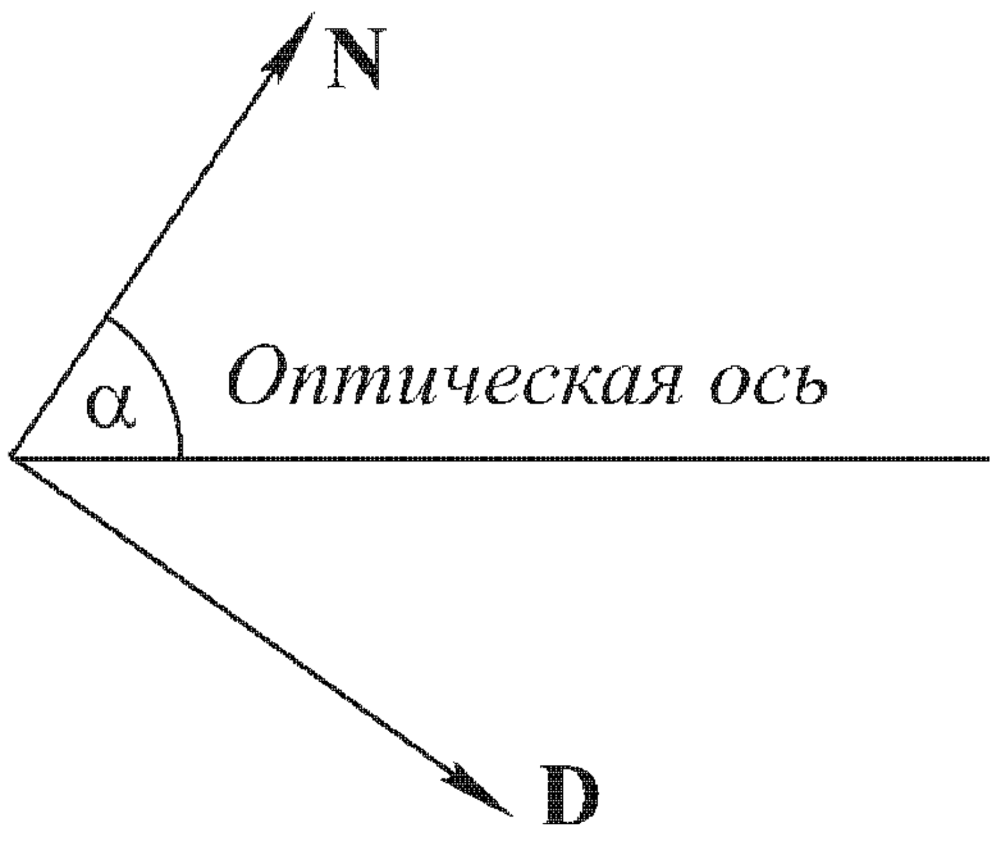


Рис. 258

Термин «оптическая ось» был введен для обозначения такой прямой, вдоль которой обе волны в кристалле распространяются с одинаковыми скоростями. Таких прямых в общем случае две (см. § 80 и 81). Поэтому в этом случае кристалл называется *оптически двуосным.* В рассматриваемом нами частном случае оптические оси совпадают между собой, сливаясь в одну прямую, а потому кристалл и называется *оптически одноосным.*

2. Так как уравнения Максвелла в кристаллах линейны и однородны, то *в общем случае волна, вступающая в кристалл из изотропной среды, разделяется внутри кристалла на две линейно поляризованные волны: обыкновенную, вектор электрической индукции которой перпендикулярен к главному сечению, и необыкновенную с вектором электрической индукции, лежащим в главном сечении.* Эти волны распространяются в кристалле в различных направлениях и с различными скоростями $v_\perp$ и $v_\|$. В направлении оптической оси скорости обеих волн совпадают, так что в этом направлении может распространяться волна любой поляризации.

К обеим волнам применимы все рассуждения, которыми мы пользовались при выводе геометрических законов отражения и преломления (см. § 64). Но в кристаллах они относятся *к волновым нормалям, а не к световым лучам.* Волновые нормали отраженной и обеих преломленных волн лежат в плоскости падения. Их направления формально подчиняются закону Снеллиуса

$$\frac{\sin\varphi}{\sin\psi_\perp} = n_\perp, \qquad \frac{\sin\varphi}{\sin\psi_\parallel} = n_\parallel, \tag{76.10}$$

где $n_\perp$ и $n_\parallel$ — *показатели преломления обыкновенной и необыкновенной волн*, т.е.

$$n_\perp = \frac{c}{v_\perp} = n_o, \quad n_\parallel = \frac{c}{v_\parallel} = \left(\frac{N_\perp^2}{\varepsilon_\parallel} + \frac{N_\parallel^2}{\varepsilon_\perp}\right)^{-1/2}. \tag{76.11}$$

Из них $n_\perp \equiv n_o$ не зависит, а $n_\parallel$ зависит от угла падения. Постоянная n_o называется *обыкновенным показателем преломления кристалла.* Когда необыкновенная волна распространяется перпендикулярно к оптической оси ($N_\perp = 1$, $N_\parallel = 0$),

$$n_\parallel = \sqrt{\varepsilon_\parallel} \equiv n_e. \tag{76.12}$$

Величину n_e называют *необыкновенным показателем преломления кристалла.* Ее нельзя смешивать с показателем преломления $n_\parallel$ необыкновенной волны. Величина n_e есть *постоянная*, а $n_\parallel$ — *функция направления распространения волны.* Обе величины совпадают только тогда, когда волна распространяется перпендикулярно к оптической оси.

3. Теперь легко понять происхождение *двойного лучепреломления.* Допустим, что плоская волна падает на плоскопараллельную пластинку из одноосного кристалла. При преломлении на первой поверхности пластинки волна внутри кристалла разделится на обыкновенную и необыкновенную. Эти волны поляризованы во взаимно перпендикулярных плоскостях и распространяются внутри пластинки *в разных направлениях и с разными скоростями. Волновые нормали обеих волн лежат в плоскости падения. Обыкновенный луч*, поскольку его направление совпадает с направлением волновой нормали, *также лежит в плоскости падения. Но необыкновенный луч, вообще говоря, выходит из этой плоскости.* (В случае двуосных кристаллов деление на обыкновенную и необыкновенную волны теряет смысл — внутри кристалла обе волны «необыкновенные». При преломлении волновые нормали обеих волн, конечно, остаются в плоскости падения, однако оба луча, вообще говоря, выходят из нее.)

Если падающая волна ограничена диафрагмой, то в пластинке получатся два пучка света, которые при достаточной толщине пластинки окажутся *разделенными пространственно.* При преломлении на второй границе пластинки из нее выйдут *два пучка света*, параллельные падающему лучу. Они будут линейно поляризованы во взаимно перпендикулярных плоскостях. Если падающий свет естественный, то всегда выйдут два пучка. Если же падающий свет линейно поляризован в плоскости главного сечения или перпендикулярно к ней, то двойного преломления не получится — из пластинки выйдет *только один пучок* с сохранением исходной поляризации.

Двойное преломление возникает и при нормальном падении света на пластинку. В этом случае преломление испытывает необыкновенный луч, хотя волновые нормали и волновые фронты не преломляются. Примером может служить кристаллическая пластинка, вырезанная параллельно оптической оси кристалла (рис. 259). Параллельный пучок света, ограниченный диафрагмой, падает нормально к поверхности пластинки. Волновые фронты, обозначенные на рисунке тонкими горизонтальными линиями, всюду параллельны той же поверхности. Обыкновенный пучок лучей o преломления не испытывает. Необыкновенный e в пластинке отклоняется вбок, но по выходе из нее снова идет в первоначальном направлении.

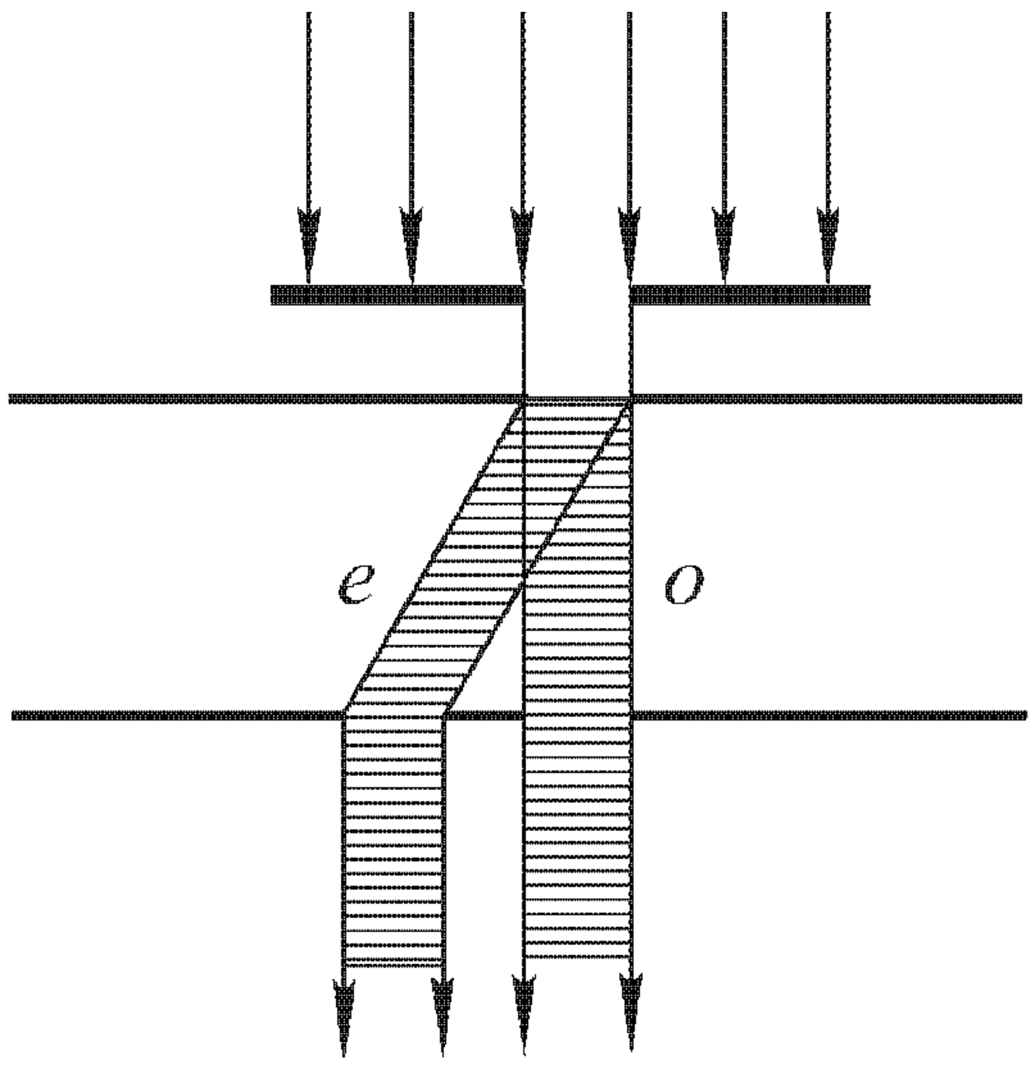


Рис. 259

4. Двойное лучепреломление было открыто в 1669 г. Бартолинусом (1625–1698) на кристаллах исландского (известкового) шпата. Гюйгенс дал объяснение этого явления, введя гипотезу, что элементарная (вторичная) волна в кристалле распадается на две волны: сферическую (обыкновенную) и эллипсоидальную (необыкновенную). Гюйгенс же открыл и поляризацию света при двойном лучепреломлении.

Исландский шпат есть разновидность углекислого кальция ($CaCO_3$). Он встречается в природе в виде довольно больших и оптически чистых кристаллов. Его обыкновенный показатель преломления $n_o = 1{,}6585$, необыкновенный $n_e = 1{,}4863$ (для желтой линии). Благодаря большому различию n_o и n_e двойное преломление в исландском шпате выражено очень отчетливо. И до сих пор кристаллы исландского шпата наиболее удобны для демонстрации двойного лучепреломления и являются наилучшим материалом для изготовления поляризационных призм и других поляризационных приборов, хотя теперь известно много естественных и искусственных кристаллов с аналогичными свойствами.

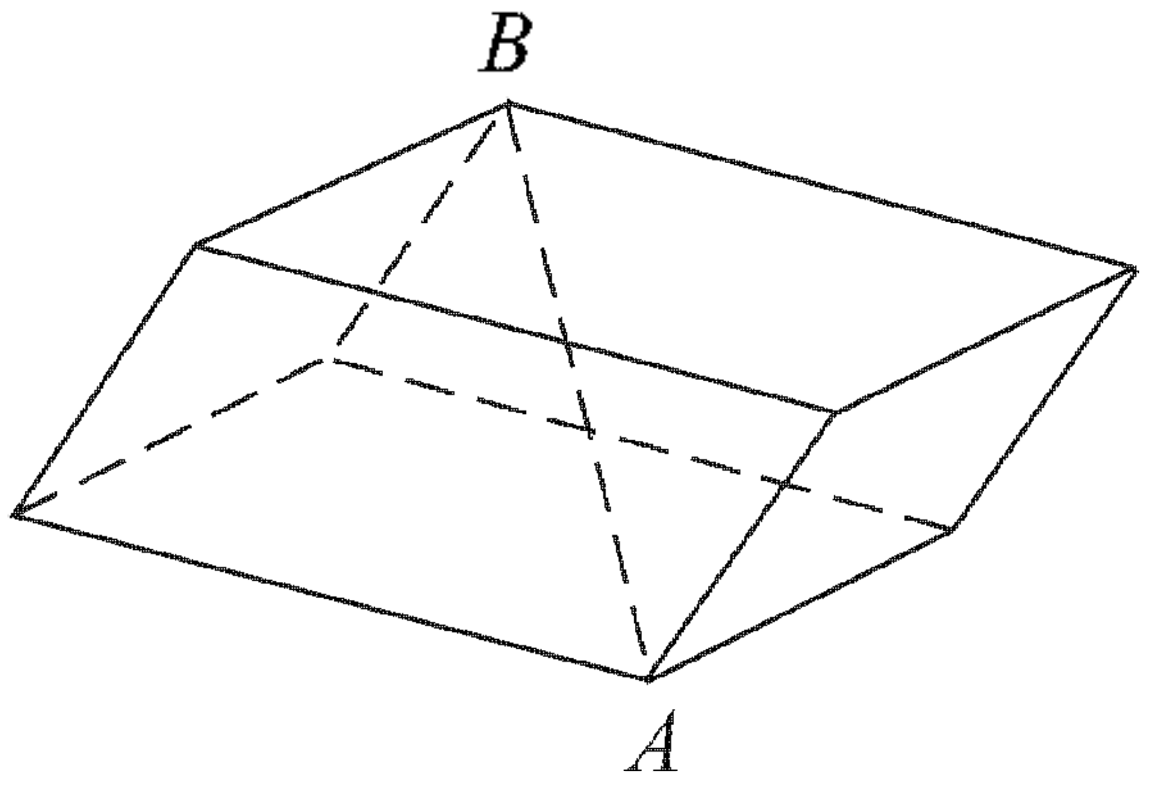


Рис. 260

Кристаллы исландского шпата принадлежат к гексагональной системе, но встречаются в различных формах. Каждый кристалл раскалыванием легко привести к форме ромбоэдра, ограниченного шестью подобными параллелограммами с углами 78°08′ и 101°52′ (рис. 260). В двух противоположных вершинах A и B сходятся стороны трех тупых углов, в остальных — стороны одного тупого и двух острых. Прямая, проходящая через точку A или B и одинаково наклоненная к ребрам, сходящимся в этих точках, а также всякая прямая, ей параллельная, есть оптическая ось кристалла. Если отшлифовать кристалл так, чтобы все ребра его имели одинаковую длину, то линия AB и будет оптической осью.

5. Для демонстрации двойного преломления берут ромбоэдр исландского шпата, вставленный в круглую вращающуюся оправу, устанавливаемую на подставке оптической скамьи. Луч света должен проходить перпендикулярно к противоположным граням ромбоэдра. Перед конденсором проекционного фонаря на той же оптической скамье устанавливают ирисовую диафрагму. В отсутствие исландского шпата длиннофокусный объектив фонаря дает изображение отверстия диафрагмы на удаленном экране. При введении непосредственно за диафрагмой исландского шпата изображение раздваивается. Уменьшая диаметр диафрагмы, можно добиться, чтобы оба изображения не накладывались друг на друга. При освещении естественным светом изображения получаются одинаково яркими. Если вращать исландский шпат вокруг главной оптической оси установки, то при правильной юстировке одно изображение остается неподвижным, а другое движется вокруг него по кругу. Когда кристалл исландского шпата делает полоборота, второе изображение совершает полный оборот. С помощью поляроида, поставленного за объективом, легко убедиться, что оба пучка света линейно поляризованы во взаимно перпендикулярных плоскостях. Неподвижное изображение дает обыкновенный, а вращающееся — необыкновенный пучок света. Увеличив диафрагму, можно частично наложить одно изображение на другое. При вращении поляроида неперекрывающиеся части изображения периодически становятся светлыми и темными: когда максимальна освещенность одной части, освещенность другой обращается в нуль. Освещенность же перекрывающейся части при вращении поляроида все время остается неизменной — эта часть освещается неполяризованным светом.

ЗАДАЧИ

1. Если пластинку исландского шпата положить на страницу печатного текста, то происходит раздвоение букв. Будет ли происходить удвоение изображения, если через ту же пластинку смотреть на удаленный предмет?

Ответ. Не будет.

2. Показатели преломления кристалла можно измерить с помощью кристалл-рефрактометра, действующего по следующему принципу. Пластинка исследуемого кристалла кладется на плоскую поверхность стеклянного полушария с очень высоким (до 2) показателем преломления N. Свет падает со стороны стеклянного полушария вдоль его радиуса и отражается от пластинки. Показатель преломления n исследуемого вещества вычисляется по предельному углу полного отражения по формуле $n = N \sin \varphi$. В случае отражения от кристалла существуют два предельных угла, соответствующих двум преломленным лучам в кристалле.

Как надо ориентировать на кристалл-рефрактометре плоскопараллельную пластинку, вырезанную произвольным образом из одноосного кристалла, чтобы получить оба главных показателя преломления кристалла n_o и n_e?

Решение. Для определения обыкновенного показателя преломления n_o пластинка, очевидно, может быть ориентирована как угодно. Чтобы получить необыкновенный показатель преломления n_e (т.е. максимальное или минимальное значение $n_\parallel$), пластинку надо ориентировать так, чтобы плоскость, проходящая через оптическую ось кристалла и нормаль к границе раздела пластинки со стеклом кристалл-рефрактометра, была перпендикулярна к плоскости падения.

Действительно, при полном отражении световое поле проникает во «вторую» среду, вообще говоря, в виде неоднородной (поверхностной) волны. Но если свет падает строго под предельным углом полного отражения, то волна во второй среде будет еще однородна. Ее волновая нормаль параллельна линии пересечения плоскости падения с плоскостью раздела сред. Повернем кристалл так, чтобы его оптическая ось стала перпендикулярна к этой линии. Тогда волна в кристалле будет распространяться перпендикулярно к оптической оси. При такой ориентировке кристалла электрический вектор необыкновенной волны будет параллелен оптической оси. Значит, на кристалл-рефрактометре в этом случае будет измерен главный показатель преломления необыкновенной волны, т.е. величина n_e.

3. При измерении показателей преломления кристаллической пластинки на кристалл-рефрактометре оказалось, что один показатель преломления сохраняет постоянное значение n_o при всех поворотах стеклянного полушария. Другой показатель преломления изменяется так, что значение n_o для него 1) максимально; 2) минимально. Определить оптический знак кристалла [1]) и ориентировку пластинки относительно оптической оси.

Ответ. Пластинка вырезана параллельно оптической оси. 1) Кристалл отрицателен. 2) Кристалл положителен.

---

[1]) Одноосный кристалл называется *положительным*, если обыкновенный показатель преломления n_o меньше необыкновенного n_e ($n_o < n_e$). В противоположном случае ($n_o > n_e$) кристалл называется *отрицательным*.

4. Как будут меняться при вращении полушария кристалл-рефрактометра оба показателя преломления пластинки, если она вырезана перпендикулярно к оптической оси?

Ответ. Оба показателя преломления будут оставаться постоянными и равными соответственно n_o и n_e.

5. Как надо вырезать призму из одноосного кристалла, чтобы методом наименьшего отклонения преломленного луча измерить обыкновенный n_o и необыкновенный n_e показатели преломления кристалла?

Ответ. Оптическая ось должна быть параллельна преломляющему ребру призмы.

6. Узкий параллельный пучок неполяризованного света падает нормально на пластинку исландского шпата, вырезанную не перпендикулярно к оптической оси. После этого пучок падает на вторую такую же пластинку, параллельную первой, а затем попадает на экран. Главные сечения пластинок образуют между собой угол 30°. Какая картина будет наблюдаться на экране?

Ответ. На экране образуются четыре пятна, интенсивности которых относятся как $1:3:1:3$. Два из них с интенсивностями $1:3$ поляризованы линейно, два других с такими же интенсивностями поляризованы также линейно, но в перпендикулярной плоскости.

7. Ветровое стекло и фары автомашины прикрыты поляроидами, главные сечения которых параллельны между собой и составляют угол 45° с горизонтом. У всех машин они должны быть повернуты в одну и ту же сторону (считая по ходу машины). Показать, что при таком устройстве водитель машины будет видеть дорогу, освещенную светом его фар, но ему не будет мешать слепящее действие фар встречных машин.

8. Определить максимальный угол $\delta_{\text{макс}}$ между направлениями луча и волновой нормали в кристалле исландского шпата.

Решение. Векторы **N**, **S**, **D**, **E** в необыкновенной волне лежат в одной плоскости, причем угол δ между **N** и **S** равен углу между **D** и **E**. Обозначим через β угол между **D** и оптической осью кристалла, а через γ — между **E** и той же осью. Тогда

$$D_{\parallel} = D\cos\beta = \varepsilon_{\parallel}E_{\parallel} = \varepsilon_{\parallel}E\cos\gamma,$$
$$D_{\perp} = D\sin\beta = \varepsilon_{\perp}E_{\perp} = \varepsilon_{\perp}E\sin\gamma.$$

Отсюда $\operatorname{tg}\beta = (\varepsilon_{\perp}/\varepsilon_{\parallel})\operatorname{tg}\gamma = (n_o/n_e)^2\operatorname{tg}\gamma$. Искомый угол $\delta = \beta-\gamma$. Поэтому

$$\operatorname{tg}\delta = \operatorname{tg}(\beta-\gamma) = \frac{n_o^2-n_e^2}{n_o^2+n_e^2\operatorname{tg}^2\beta}. \tag{76.13}$$

Максимум достигается, когда $\operatorname{tg}\beta = n_o/n_e = 1{,}1157$, т.е. при $\beta = 48°7'53''$. Максимальный угол $\delta_{\text{макс}}$ определяется формулой

$$\operatorname{tg}\delta_{\text{макс}} = \frac{n_o^2-n_e^2}{2n_on_e} = 0{,}1097. \tag{76.14}$$

Из нее находим $\delta_{\text{макс}} = 6°15'46''$.

9. Определить показатель преломления $n_{\parallel}$ необыкновенной волны в ромбоэдре исландского шпата, если волновая нормаль параллельна боковому ребру кристалла. Для исландского шпата угол между боковым ребром и оптической осью равен $\alpha = 64°$. Найти для этого случая угол δ между направлениями луча и волновой нормали в кристалле.

Ответ. $n_{\parallel} = \sqrt{(\sin\alpha/n_e)^2+(\cos\alpha/n_o)^2} = 1{,}515$; $\delta = 4°36'20''$.

§ 77. Поляризационные устройства

1. Двупреломляющие кристаллы (лучше всего исландского шпата) можно использовать для получения поляризованного света. Для этого в опыте типа рис. 259 достаточно задержать один из двух линейно поляризованных пучков света, вышедших из кристалла. Однако гораздо удобнее пользоваться не простыми кристаллами, а их комбинациями, называемыми *поляризационными призмами.* Обычно для изготовления поляризационных призм применяется исландский шпат, иногда кварц и натронная селитра, что значительно удешевляет стоимость поляризационных приборов.

Поляризационная призма состоит из двух или более трехгранных призм из одноосного кристалла с одинаковой или различной ориентацией оптических осей, склеенных между собой прозрачными веществами или разделенных воздушной прослойкой. Для склейки применяются канадский бальзам ($n_D = 1{,}550$), льняное масло ($n_D = 1{,}49$), глицерин ($n_D = 1{,}474$) и другие материалы. Для работы в ультрафиолетовой части спектра применяют призмы, склеенные глицерином, касторовым маслом, а также призмы с воздушной прослойкой.

Различают *однолучевые поляризационные призмы*, из которых выходит только один пучок линейно поляризованного света, и *двухлучевые поляризационные призмы*, из которых выходят два пучка света, поляризованных во взаимно перпендикулярных плоскостях. Призмы первого типа действуют по принципу *полного отражения.* Падающий пучок естественного света, проникая в призму, расщепляется в ней на два пучка, поляризованных линейно во взаимно перпендикулярных плоскостях. Один из этих пучков претерпевает полное отражение на границе склейки призм и отклоняется вбок, другой проходит прямо и используется в дальнейшем.

2. Первая поляризационная призма была изобретена в 1828 г. шотландским физиком Николем (1768–1851). Ее сокращенно называют *николем.* То же название часто применяют и для других призм, действующих по тому же принципу. Для изготовления призмы Николя у продолговатого ромбоэдра, полученного скалыванием из куска исландского шпата, сошлифовывают основания так, чтобы новые основания составляли с боковыми ребрами угол 68° (вместо 71° у естественного кристалла). Затем кристалл разрезают вдоль плоскости, перпендикулярной к новым основаниям и к главному сечению кристалла (последнее нормально к этим основаниям и проходит через оптическую ось кристалла). Отполировав плоскости разреза, оба куска склеивают в прежнем положении тонким слоем канадского бальзама.

Сечение призмы Николя плоскостью главного сечения показано на рис. 261. Двойная стрелка, наклоненная под углом 64° к длинному ребру, указывает направление оптической оси. Такое обозначение применяется в дальнейшем и для других призм. Луч света, падая на искусственное основание кристалла, разделяется внутри кристалла на обыкновенный AO и необыкновенный AE. Показатель преломления канадского бальзама (1,550) имеет промежуточное значение между обыкновенным ($n_o = 1{,}658$) и необыкновенным ($n_e = 1{,}486$) показателями преломления исландского шпата. Углы в призме Николя рассчитаны так, чтобы необыкновенный луч прошел через слой канадского бальзама, а обыкновенный претерпел на нем полное отражение и поглотился зачерненной боковой гранью. (В больших призмах во избежание их нагревания обыкновенный луч выводится наружу специальной призмочкой, наклеенной на боковую грань и изображенной на рис. 261 штриховой линией.) В результате свет, вышедший из призмы, окажется линейно поляризованным.

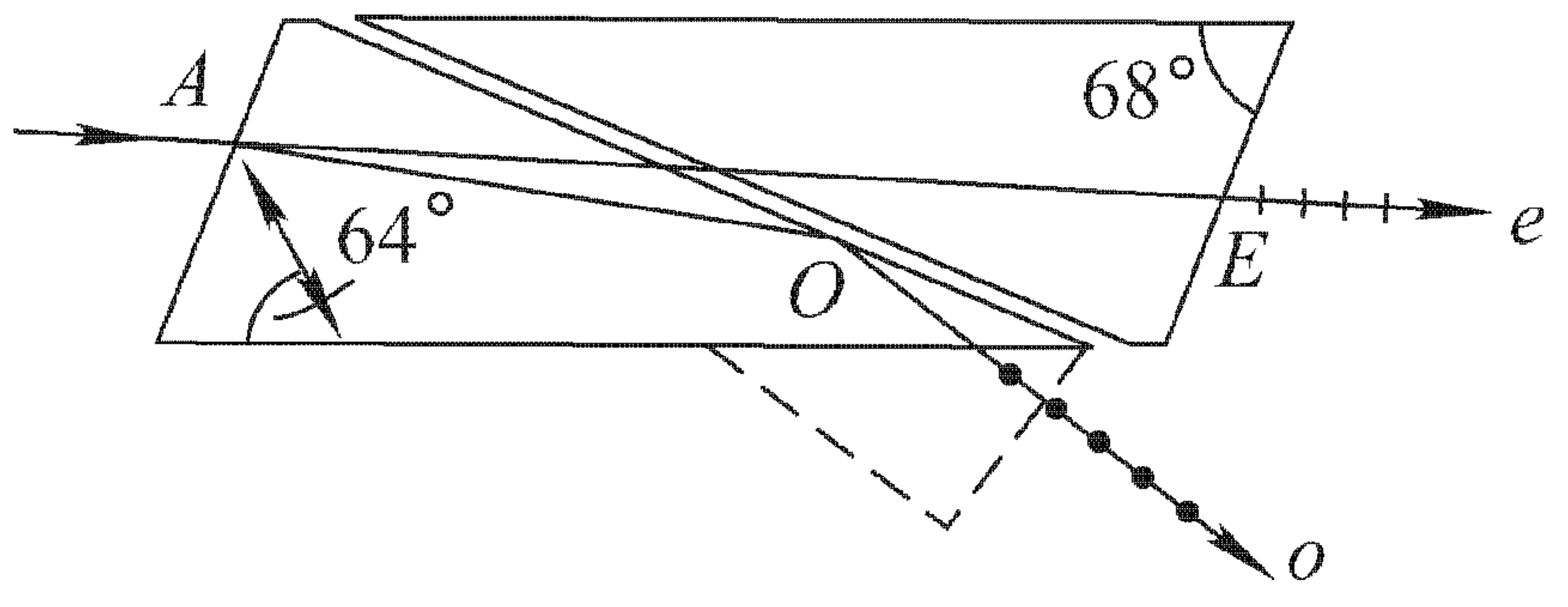


Рис. 261

В пучке лучей, падающих на входную грань призмы Николя, обычно встречаются лучи разных направлений. Если луч падает на плоскость разреза слишком наклонно, то претерпит полное отражение не только обыкновенный луч, но и необыкновенный. Если же наклон недостаточен, то через призму пройдет и обыкновенный луч. Через призму может пройти один только необыкновенный луч тогда и только тогда, когда углы падения на входную грань лежат в определенных пределах. Разность углов наклона между крайними лучами падающего пучка, удовлетворяющими этому условию, определяет так называемую *апертуру полной поляризации призмы*. Для призмы Николя она составляет 29°, а отношение длины призмы к ее ширине — 3,28 (см. задачу 1 к этому параграфу).

В *призме Фуко* (рис. 262), устроенной как и николь, канадский бальзам заменен тонким слоем воздуха. Благодаря этому она может применяться для ультрафиолета, тогда как призма Николя непригодна для этой цели, так как канадский бальзам поглощает ультрафиолетовый свет. От николя призма Фуко отличается еще тем, что показатель преломления воздушной прослойки меньше обоих показателей n_o и n_e, а не равен промежуточному значению между ними, как это было в случае канадского бальзама. Поэтому полное отражение от плоскости разреза в призме Фуко осуществляется под меньшими углами, а сама призма Фуко (при заданном поперечном сечении) гораздо короче и, следовательно, дешевле николя. Зато апертура полной поляризации ее составляет всего 8°.

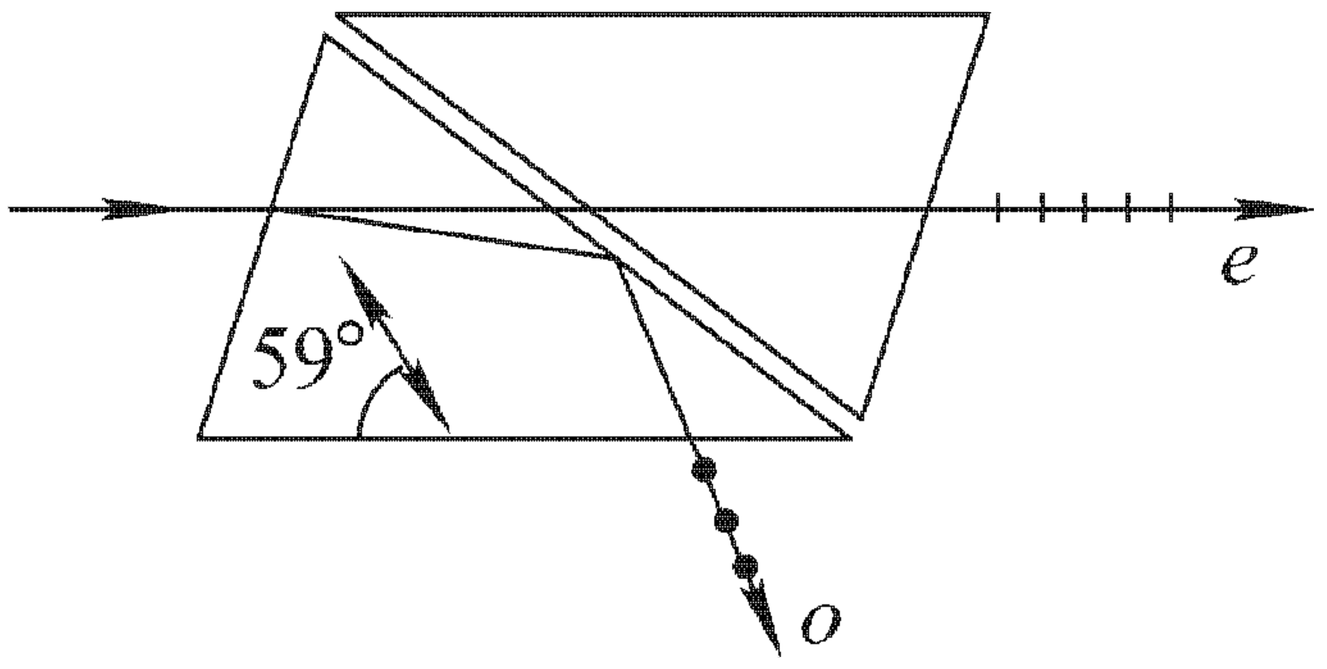


Рис. 262

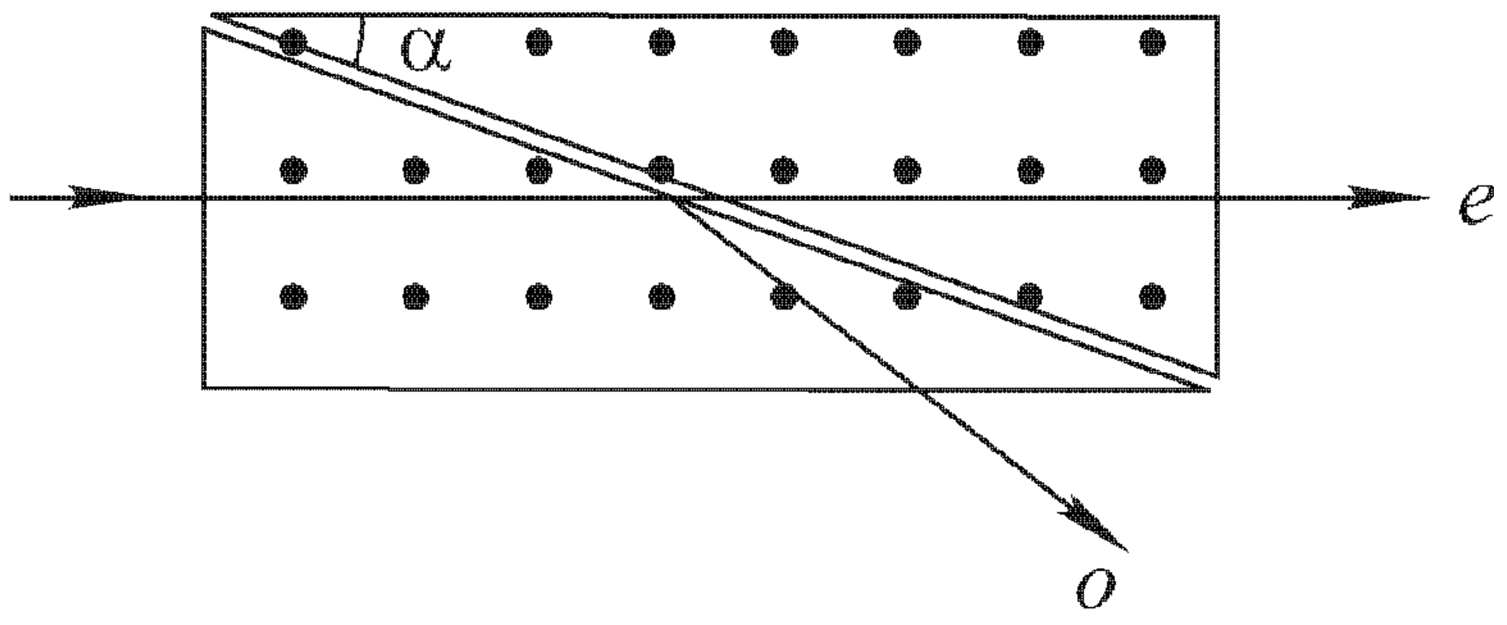


Рис. 263

3. Призмы Николя и Фуко имеют скошенные основания. Это вызывает параллельное боковое смещение падающего луча при прохождении его через призму. Следствием этого является кругообразное перемещение выходящего луча при вращении призмы вокруг ее оси. От этого недостатка свободны призмы, имеющие форму прямоугольного параллелепипеда: Глазебрука (рис. 263), Глана, Глана–Томпсона, Гартнака–Празмовского, Франка–Риттера (рис. 264), Аренса (рис. 265) и др. Сечение склеенной поляризационной призмы Глазебрука показано на

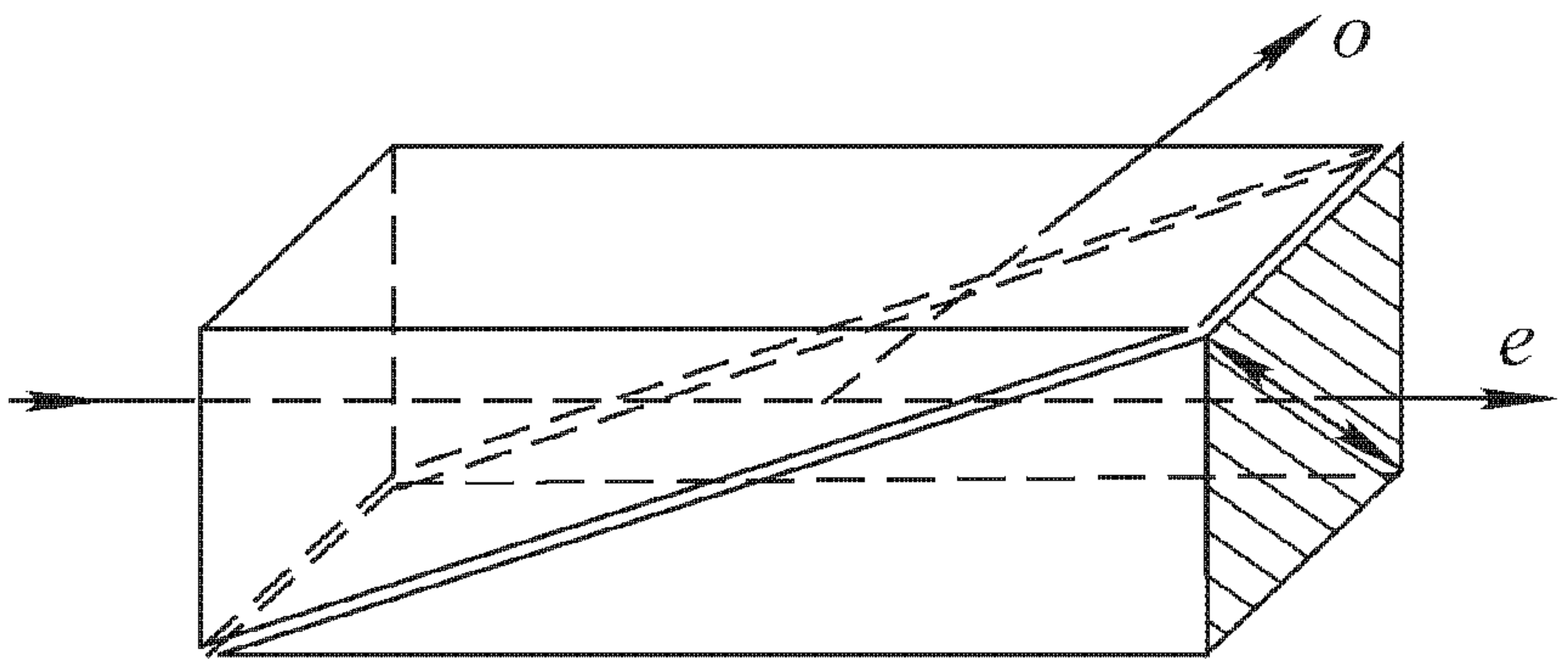


Рис. 264

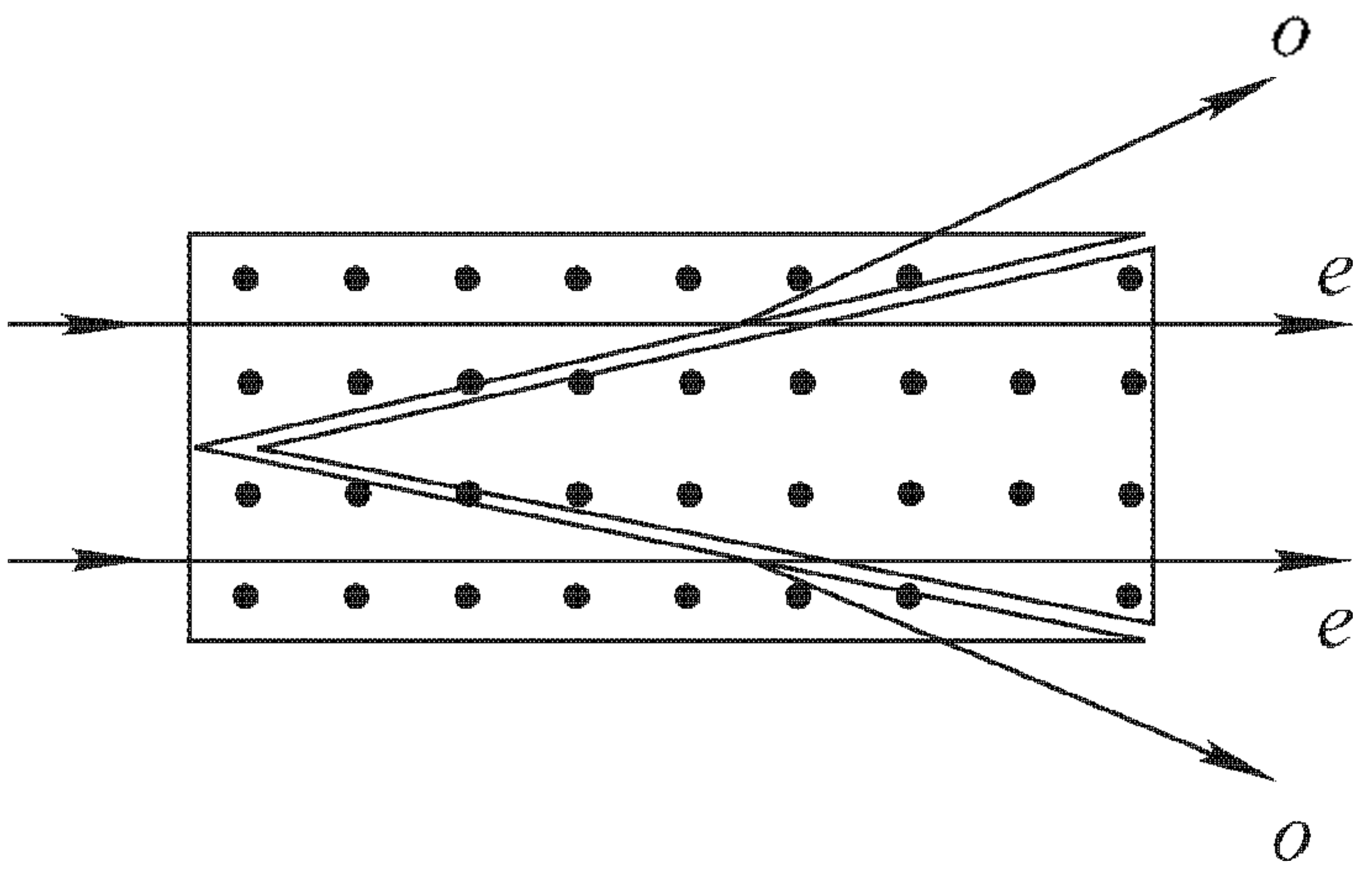


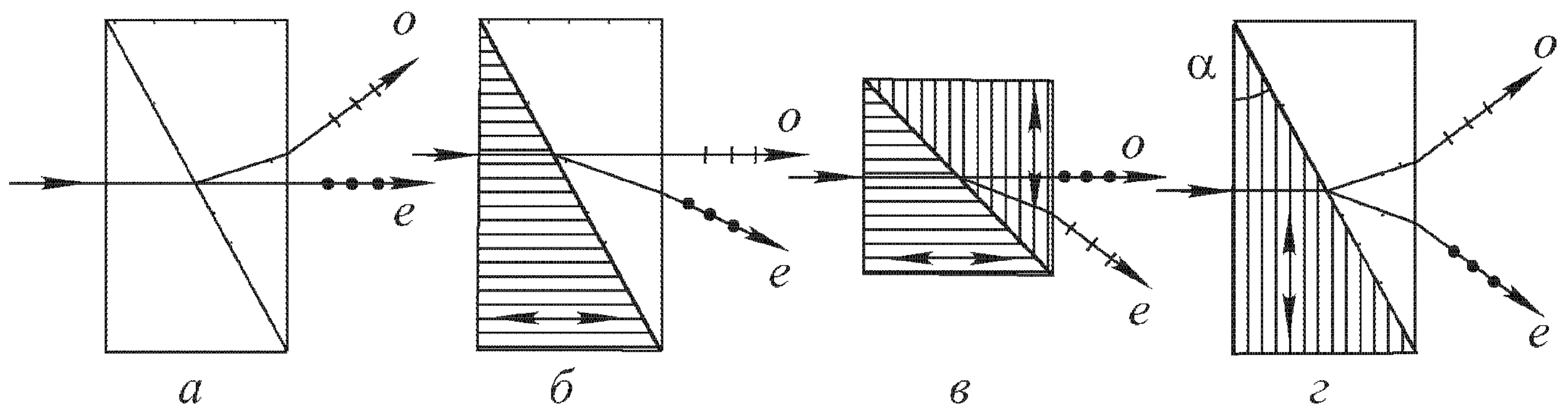
Рис. 265

рис. 263. Оптическая ось перпендикулярна к плоскости рисунка, что отмечено точками. Аналогично устроены призмы Глана, Глана–Томпсона и Гартнака–Празмовского. Их отличия от призмы Глазебрука заключаются в следующем. Призма Глана — воздушная. Призмы Глана–Томпсона имеют другие параметры. В призме Гартнака–Празмовского оптическая ось нормальна к плоскости распила, а в призме Франка–Риттера (рис. 264) она параллельна одной из диагоналей основания призмы. Призма Аренса — тройная призма, устройство и действие которой понятно из рис. 265. Параметры поляризационных призм указаны в табл. 8.

Таблица 8

Вид призмы	Апертурный угол полной поляризации	Угол плоскости разреза с боковой гранью	Отношение длины к ширине
Призма Николя	29°	22°	3,28
Укороченная призма Николя, склеенная:			
а) канадским бальзамом,	13°	25°	2,83
б) копайским бальзамом	24°	25°	2,83
Призма Фуко	8°	40°	1,53
Призма Глазебрука, склеенная:			
а) канадским бальзамом,	35°	12,1°	4,65
б) льняным маслом,	42°	14°	4,02
в) глицерином	32,1°	17,3°	3,2
Призма Глана с воздушной прослойкой	7,9°	50,3°	0,83
Призма Глана–Томпсона, склеенная:			
а) канадским бальзамом,	27,5°	13,5°	4,15
б) льняным маслом	41°	13,5°	4,15
Призма Гартнака–Празмовского	41,9°	13,9°	4,04
Призма Франка–Риттера, склеенная канадским бальзамом	19°	17,5°	3,17
Призма Аренса, склеенная канадским бальзамом	35°	12,1°	2,32

4. Различные типы *двухлучевых поляризационных призм* представлены на рис. 266. Первая призма состоит из комбинации стеклянной призмы с кристаллической из исландского шпата, оптическая ось которой параллельна преломляющему ребру. Призмы соприкасаются или склеиваются, как показано на рис. 266 *а*. Показатель преломления стекла (1,49) почти точно совпадает с необыкновенным показателем преломления n_e



а *б* *в* *г*

Рис. 266

исландского шпата. Падающий пучок неполяризованного света в кристаллической призме разделяется на обыкновенный и необыкновенный. Необыкновенный проходит без преломления. Обыкновенный сильно отклоняется к основанию кристаллической призмы в результате двукратного преломления на ее гранях. Аналогично действуют *призмы Рошона* (рис. 266 *б*) и *Сенармона* (рис. 266 *в*). В *призме Волластона* (рис. 266 *г*) выходящие лучи линейно поляризованного света отклоняются в разные стороны.

5. У многих кристаллов поглощение света зависит от направления электрического вектора в световой волне. Это явление также используется для получения линейно поляризованного света в так называемых *дихроичных пластинках*. К ним относятся, например, пластинки турмалина и поляроиды. В турмалине обыкновенный луч поглощается сильнее необыкновенного. Поэтому после прохождения через пластинку турмалина естественный свет становится частично поляризованным в плоскости главного сечения. Если пластинка достаточно толстая (около 1 мм), то в области видимого света обыкновенный луч поглощается практически полностью, так что прошедший свет окажется полностью линейно поляризованным. Для некоторых участков видимого спектра и необыкновенный луч испытывает заметное поглощение. Турмалин при надлежащей толщине действует не только как *поляризатор*, но и как *светофильтр*, пропускающий преимущественно желто-зеленую область спектра. Это — существенный недостаток турмалина как поляризатора. Зато апертура полной поляризации у него очень велика и ограничена одними только конструктивными условиями.

Заметим еще, что с изменением направления волновой нормали меняется и направление электрического вектора в обеих линейно поляризованных волнах, на которые разделяется естественный свет, распространяющийся вдоль этой нормали. Вследствие этого меняется и поглощение таких волн, которое к тому же обычно зависит от длины волны. Это приводит к зависимости окраски кристалла от направления распространения света. Такое явление называется *дихроизмом* (двуцветностью) или, лучше, *плеохроизмом* (многоцветностью) и в большей или меньшей степени присуще большинству двупреломляющих кристаллов.

ЗАДАЧИ

1. В поляризационной призме Николя плоскость разреза BC (рис. 267) перпендикулярна к основаниям AB и CD. Разрез произведен по диагонали призмы. Определить предельные углы полного отражения от слоя канадского бальзама для обыкновенной и необыкновенной волн. Найти соответствующие им углы падения на основание AB и апертуру полной поляризации призмы. Найти показатель преломления $n_{\|}$ необыкновенной волны в кристалле. Найти также отношение длины l призмы к ширине h ее поперечного сечения. Вычисления произвести для двух случаев: 1) основания призмы сошлифованы ($\gamma = 68°$); 2) для естественного кристалла ($\gamma = 71°$).

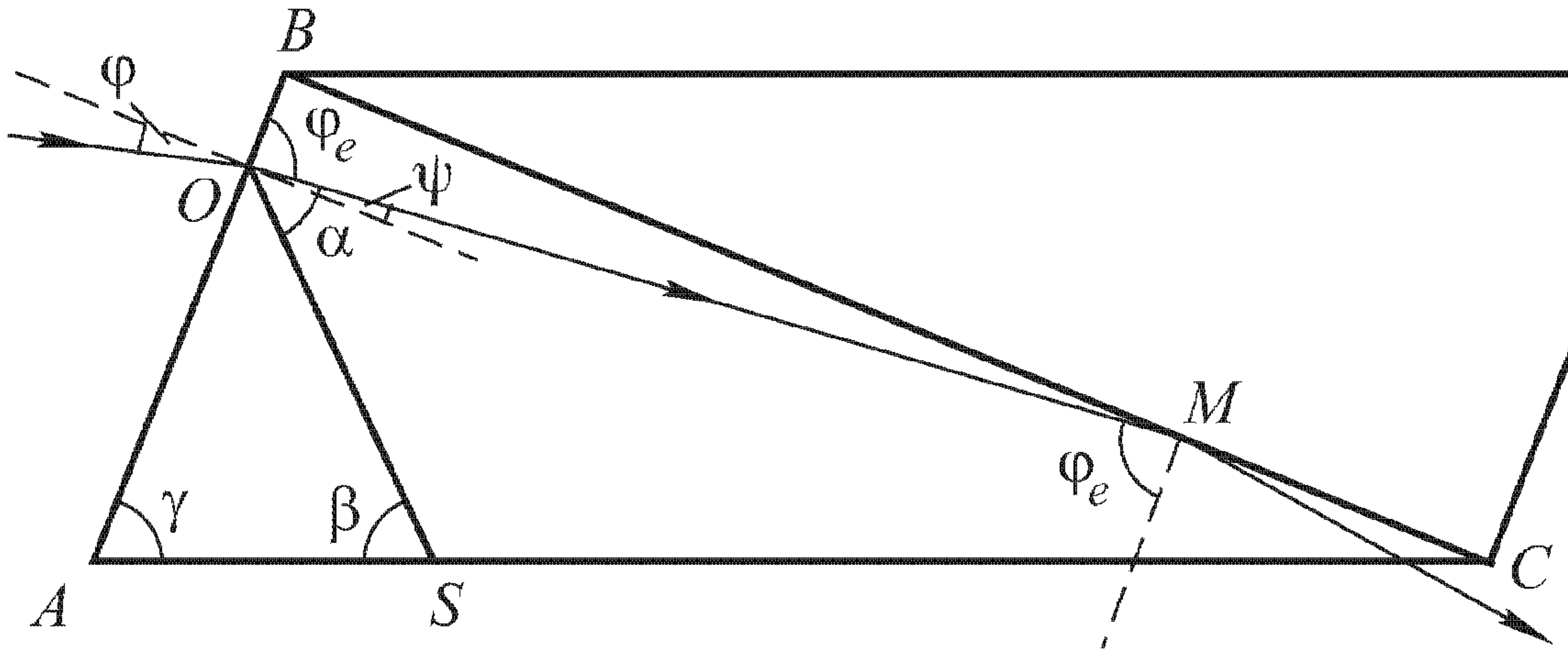


Рис. 267

Решение. Проведем расчет для необыкновенной волны в предположении, что $\gamma = 68°$. Пусть $\beta = 64°$ — угол между оптической осью кристалла и продольным ребром AC, OM — волновая нормаль для предельного случая полного отражения, φ_e — соответствующий угол падения на слой канадского бальзама BC. Тогда $\sin\varphi_e = n/n_{\|}$, где n — показатель преломления канадского бальзама. Показатель преломления $n_{\|}$ необыкновенной волны в кристалле для рассматриваемого направления волновой нормали определяется уравнением

$$\frac{1}{n_{\|}^2} = \frac{\sin^2\alpha}{n_e^2} + \frac{\cos^2\alpha}{n_o^2}.$$

Как видно из рис. 267, $\alpha = \beta + \gamma - \varphi_e$. Подставив числовые значения в предыдущее соотношение, придем к уравнению

$$\operatorname{tg}^2\varphi - 7{,}046608\operatorname{tg}\varphi - 32{,}84593 = 0.$$

Из него находим

$$\operatorname{tg}\varphi = 10{,}250829, \quad \varphi = 84{,}422825°, \quad n_{\|} = n/\sin\varphi_e = 1{,}5573726.$$

Угол преломления на основании AB равен $\psi = \pi/2 - \varphi_e = 5{,}5771175°$. По закону Снеллиуса находим угол падения

$$\sin\varphi = n_{\|}\sin\psi = 0{,}1513532, \quad \varphi = 8{,}705349°.$$

Чтобы необыкновенная волна не испытывала полного отражения, угол падения φ должен удовлетворять условию $\varphi > 8{,}705349°$.

Аналогичный расчет для обыкновенной волны, разумеется, проще, так как показатель преломления такой волны не зависит от направления ее распространения. Для предельного угла полного отражения в этом случае получаем $\varphi_o = 69{,}20573°$, а для соответствующего угла падения $\varphi = 36{,}05854°$. Чтобы обыкновенная волна испытывала полное отражение, угол падения должен быть меньше этого значения. При угле падения, удовлетворяющем условию $8{,}705349° < \varphi < 36{,}05854°$, прошедший свет будет поляризован полностью. Апертура полной поляризации $\varepsilon = 36{,}06 - 8{,}71 \approx 27{,}35°$. (При вычислениях использовано неточное значение угла β. Десятичные знаки сохранялись для того, чтобы не потерять точность во время расчета.)

Для отношения l/h получаем

$$\frac{l}{h} = \frac{1}{\sin\gamma\cos\gamma} = \frac{2}{\sin 2\gamma} = 2{,}88.$$

Практически призму Николя распиливают не по диагонали, а как указано на рис. 261, чтобы легче монтировать ее. В этом случае отношение l/h несколько больше.

Рассмотрим теперь случай естественного кристалла, когда $\gamma = 71°$. Вычисления для обыкновенной волны остаются без всяких изменений. Вычисления же для необыкновенной волны сильно упрощаются благодаря случайному соотношению: $\beta + \gamma = 135°$, и следовательно, $\sin(\beta+\gamma) = -\cos(\beta+\gamma)$. Приводим окончательные результаты: $n_{\parallel} = 1{,}564955$, $\varphi_e = 82{,}0728°$, $\varphi = 12{,}46454°$, $\varepsilon = 23{,}59°$, $l/h = 3{,}25$. По сравнению с предыдущим случаем параметры поляризационной призмы ухудшились.

2. Призма Волластона (рис. 266 *г*) сделана из исландского шпата. Угол $\alpha = 15°$. На какой угол δ будут разведены обыкновенный и необыкновенный лучи?

Ответ. $\delta = 5°17'$.

3. Две поляризационные призмы с воздушной прослойкой изготовлены из исландского шпата. В одной призме оптическая ось перпендикулярна, в другой — параллельна плоскости падения (рис. 268). Опишите действие каждой призмы. Как будет поляризован проходящий свет? Какая призма будет пропускать больше света? В каких пределах должен быть заключен угол α, чтобы из призмы выходил линейно поляризованный свет? Свет падает на грань призмы перпендикулярно.

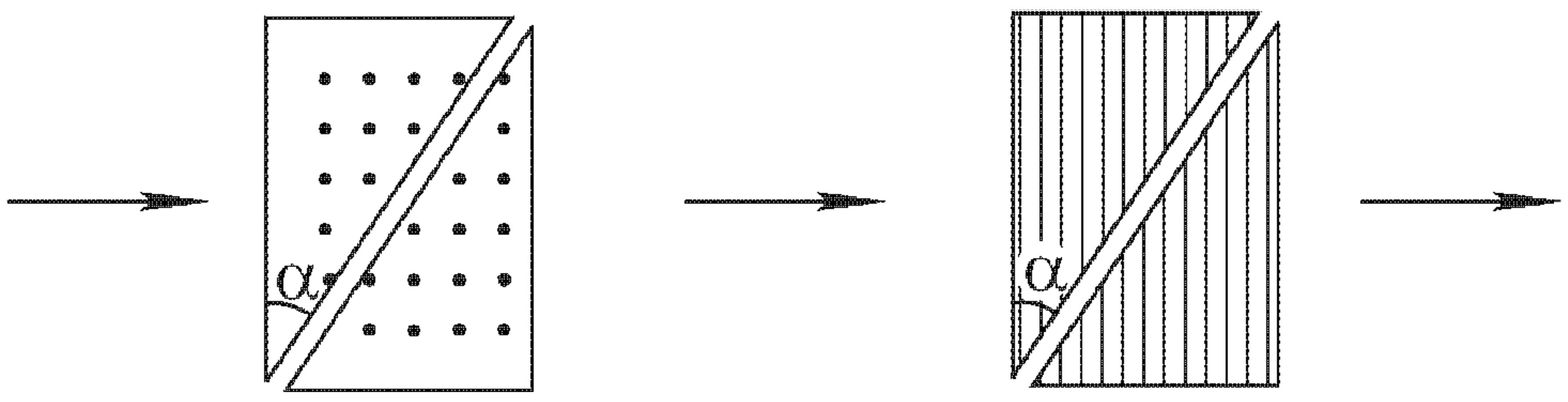


Рис. 268

Ответ. Через призмы проходит необыкновенный луч. Вторая призма пропускает больше света. В обоих случаях должно выполняться условие $1/n_o < \sin\alpha < 1/n_e$, т.е. $37°6' < \alpha < 42°18'$.

4. В призме Глазебрука (рис. 263) склеивающим веществом является либо канадский бальзам ($n = 1{,}550$), либо льняное масло ($n = 1{,}494$). Найти наивыгоднейшую апертуру полной поляризации призмы и соответствующее ей отношение длины a к ширине b для обоих склеивающих материалов. Пучок падающих лучей, дающих полную поляризацию, должен быть симметричен относительно оси, параллельной боковому ребру призмы.

Решение. В рассматриваемом случае электрический вектор необыкновенной волны всегда параллелен оптической оси, а обыкновенной — перпендикулярен к ней. Поэтому направления лучей и волновых нормалей совпадают между собой. Необыкновенный луч не может претерпеть полное отражение от плоскости разреза призмы, так как n_e меньше показателя преломления склеивающего вещества. Наибольший наклон необыкновенного луча, идущего сверху вниз направо (рис. 269), определяется направлением диагонали BC. Угол преломления для такого луча $\psi_e = \arcsin\left(1/\sqrt{1+(a/b)^2}\right)$, а угол падения

$$\varphi = \arcsin\frac{n_e}{\sqrt{1+(a/b)^2}}. \tag{77.1}$$

Если угол падения превышает это значение, то необыкновенный луч на плоскость разреза не попадет, а поглотится боковой гранью AC. Подберем теперь отношение a/b так, чтобы обыкновенный луч SO, падающий под углом φ в направлении снизу вверх направо, встретил плоскость разреза BC под предельным углом полного отражения $\beta = \arcsin(n/n_o)$. Тогда всякий обыкновенный луч, падающий снизу вверх под меньшим углом, а также приходящий сверху вниз полностью отразится от плоскости разреза BC. Апертура полной поляризации будет $\varepsilon = 2\varphi$. Как видно из рис. 269, $\beta = \frac{\pi}{2} - (\psi_e + \psi_o)$, и следовательно, $\cos(\psi_e + \psi_o) = n/n_o$. Добавив сюда закон преломления $\sin\varphi = n_o\sin\psi_o = n_e\sin\psi_e$, получим полную систему уравнений для нахождения искомых величин: отношения a/b и апертуры 2φ. Из нее находим удобные расчетные формулы:

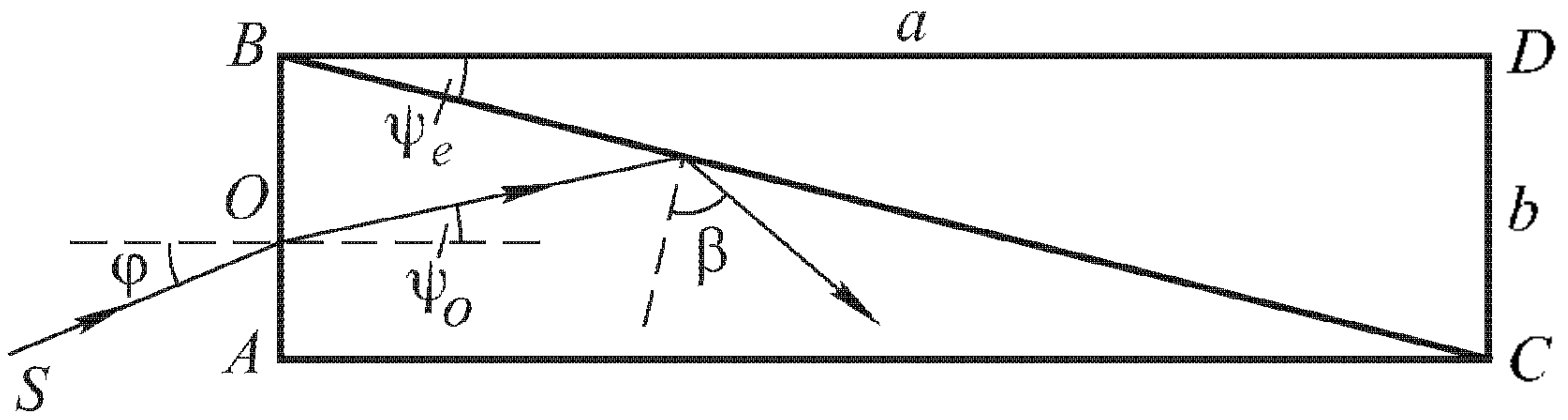


Рис. 269

$$(n_o^2 - n^2)\left(\frac{a^2}{b^2}\right)^2 + \left[n_o^2 - n_e^2 - 2n(n+n_e)\right]\frac{a^2}{b^2} - (n+n_e)^2 = 0, \tag{77.2}$$

$$\sin\frac{\varepsilon}{2} = \frac{n_e}{\sqrt{1+(a/b)^2}}. \tag{77.3}$$

Они дают для канадского бальзама $a/b = 5{,}32$, $\varepsilon = 31^\circ 51'$, для льняного масла $a/b = 4{,}25$, $\varepsilon = 39^\circ 50'$.

5. Решить предыдущую задачу в предположении, что слой склеивающего вещества заменен прослойкой воздуха.

Решение. Для того чтобы обыкновенный луч, падающий на грань AB под углом φ снизу вверх направо (рис. 270), испытал полное отражение от плоскости разреза BC, необходимо, чтобы угол падения на эту плоскость превосходил предельный угол полного отражения:

$$\alpha - \psi_o \geqslant \beta_o, \quad \sin\beta_o = 1/n_o. \tag{77.4}$$

Для того чтобы необыкновенный луч, падающий на грань AB под углом φ' сверху вниз направо, прошел через плоскость разреза BC, необходимо, чтобы он падал на BC под углом, не превосходящим предельного угла полного отражения:

$$\alpha + \psi'_e \leqslant \beta_e, \quad \sin\beta_e = 1/n_e. \tag{77.5}$$

Наибольшей апертуре призмы соответствуют в (77.4) и (77.5) знаки равенства. В этом случае угол $\varphi + \varphi'$ равен апертуре ε полной поляризации призмы. Потребовав дополнительно $\varphi = \varphi'$, будем иметь

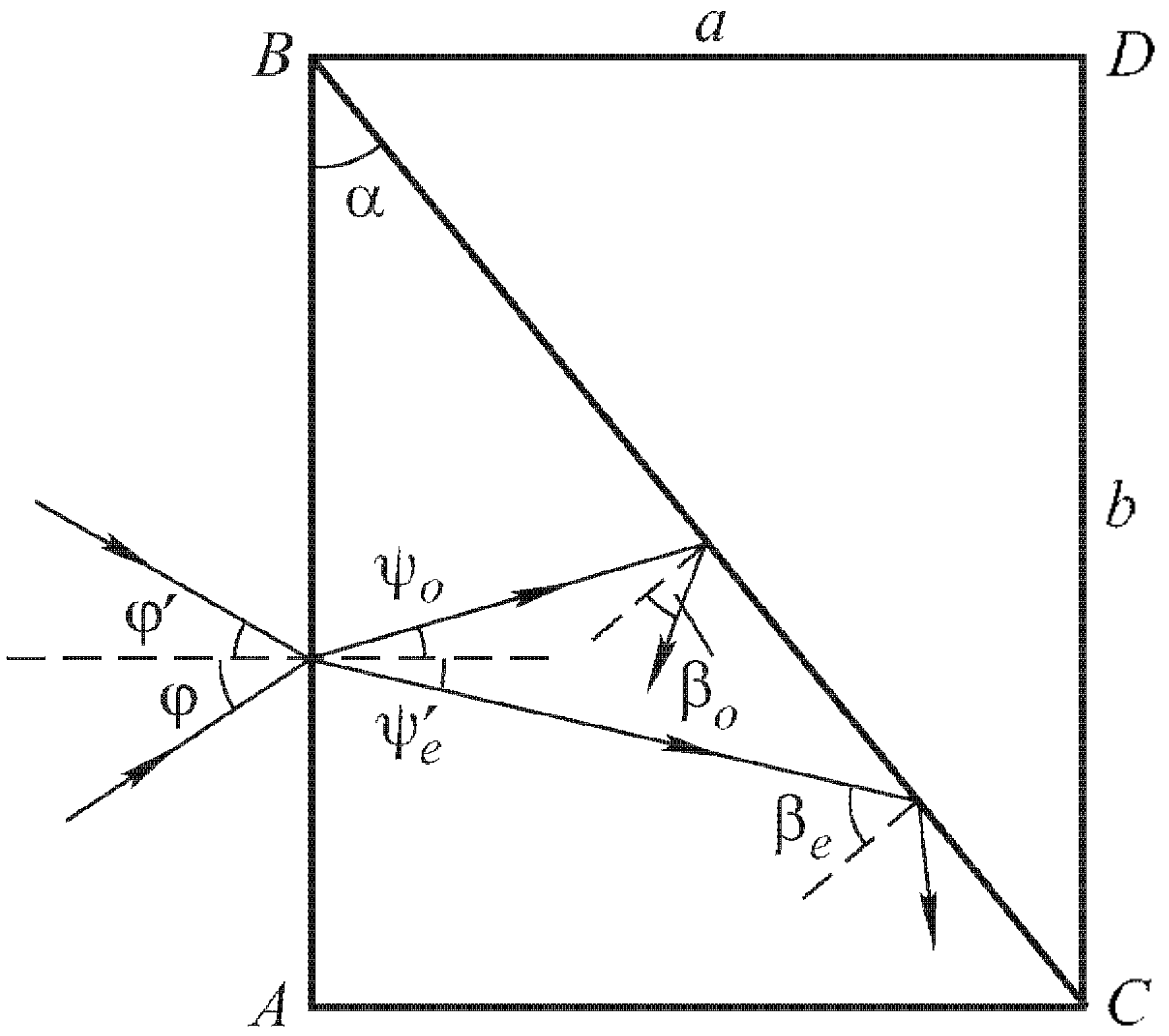


Рис. 270

$$\alpha - \psi_o = \beta_o, \quad \alpha + \psi_e' = \beta_e, \quad (77.6)$$

причем

$$\sin\varphi = n_o \sin\psi_o = n_e \sin\psi_e', \quad (77.7)$$

$$\operatorname{tg}\alpha = a/b. \quad (77.8)$$

Отсюда могут быть найдены все интересующие нас величины. Из (77.6) получаем

$$\psi_o + \psi_e' = \beta_e - \beta_o = 5^\circ 12'. \quad (77.9)$$

Таким образом, углы ψ_o, ψ_e', φ малы и их синусы можно заменить самими углами. Это дает для апертуры

$$\varepsilon = 2\varphi = \frac{2n_o n_e}{n_o + n_e}(\beta_e - \beta_o) = 8^\circ 9' \quad (77.10)$$

и для отношения сторон

$$\frac{a}{b} = \operatorname{tg}(\beta_o + \psi_o) = \operatorname{tg}\left(\beta_o + \frac{\varphi}{n_o}\right) = 0{,}826. \quad (77.11)$$

§ 78. Анализ поляризованного света

1. В линейной поляризации света можно убедиться с помощью поляризационной призмы (николя) или любого поляризатора, способного давать полностью линейно поляризованный свет. (С целью сокращения всякий поляризатор в дальнейшем называется *николем.*) Для этого николь ставят на пути исследуемого света. При вращении николя вокруг направления луча интенсивность проходящего света, вообще говоря, будет изменяться. Если при некотором положении николя проходящий свет полностью гасится, то падающий свет был поляризован *линейно.*

Если падающий свет естественный или поляризован по кругу, то при вращении николя интенсивность проходящего света меняться не будет. Для отличия одного случая от другого применяется *пластинка в четверть волны* (короче, $\lambda/4$) или *компенсатор.* Пластинка в четверть волны есть кристаллическая пластинка, которая вносит дополнительную разность фаз в $\pi/2$ между проходящими через нее лучами, поляризованными во взаимно перпендикулярных плоскостях. Эти плоскости определяют в плоскости пластинки два направления, называемые *главными направлениями пластинки.* Обычно пластинка

$\lambda/4$ вырезается из одноосного кристалла (например, кварца) параллельно его оптической оси. Тогда дополнительная разность фаз в $\pi/2$ вносится между обыкновенным и необыкновенным лучами. Но пластинку $\lambda/4$ можно изготовить и из двуосного кристалла, например слюды. В дальнейшем для определенности предполагается, что пластинка $\lambda/4$ вырезана из одноосного кристалла. В свете, поляризованном по кругу, разность фаз между любыми двумя взаимно перпендикулярными колебаниями равна $\pm\pi/2$. Если на пути такого света поставить пластинку $\lambda/4$, то она внесет дополнительную разность фаз $\pm\pi/2$. Результирующая разность фаз получится 0 или π, и свет станет поляризованным *линейно*. Его можно полностью погасить поворотом николя. Если же падающий свет естественный, то он останется таковым и после прохождения через пластинку $\lambda/4$. В этом случае гашения не будет.

2. Для отличия правой круговой поляризации от левой можно воспользоваться тем же приспособлением. Допустим, что свет распространяется вдоль оси Z по направлению к читателю, а оптическая ось пластинки $\lambda/4$ ориентирована вдоль оси Y. Тогда вектор $\mathbf{E}$ в необыкновенной волне будет параллелен оси Y, а в обыкновенной — оси X. Для определенности допустим, что поляризация правая (рис. 271), т.е. вращение вектора $\mathbf{E}$ совершается по часовой стрелке. В этом случае перед пластинкой необыкновенная волна опережает по фазе обыкновенную на $\pi/2$ (см. § 62, п. 4). Если z — толщина пластинки, то после прохождения через нее между этими волнами возникнет дополнительная разность фаз

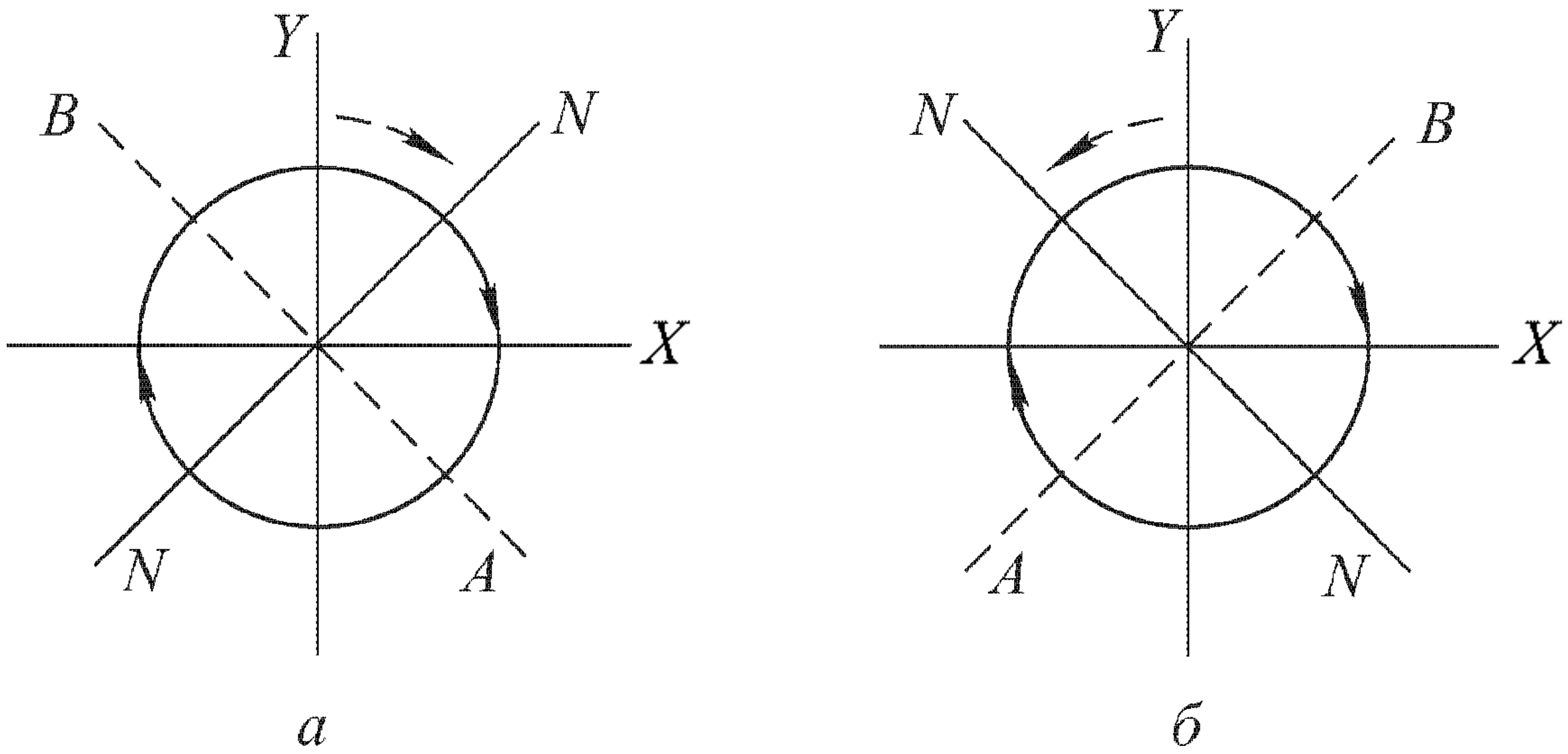


Рис. 271

$$\Delta_{\text{доп}} = (\omega t - kn_e z) - (\omega t - kn_o z) = k(n_o - n_e)z,$$

где k — волновое число в вакууме. Для отрицательного кристалла ($n_o > n_e$, исландский шпат) $\Delta_{\text{доп}} = +\pi/2$, для положительного ($n_o < n_e$, кварц) $\Delta_{\text{доп}} = -\pi/2$. В первом случае полная разность хода равна $\Delta = \pi$, и плоскость колебаний прошедшего света на рис. 271 *а* будет направлена вдоль штриховой биссектрисы AB. Если первоначально главное сечение николя проходило через оптическую ось Y пластинки $\lambda/4$, то его надо повернуть в положение NN в направлении штриховой стрелки, т.е. *в ту же сторону*, куда вращается вектор **E** в падающей волне, чтобы полностью погасить проходящий свет. Во втором случае $\Delta = 0$ (рис. 271 *б*) и николь надо повернуть в противоположную сторону, т.е. *против направления вращения вектора* **E** в падающей волне. Если поляризация падающей волны *левая*, то все вращения николя надо производить в *противоположную сторону.*

Результаты можно резюмировать в виде следующего правила. *Пусть в исходном положении плоскость главного сечения николя проходит через оптическую ось пластинки $\lambda/4$. Положительным считается вращение николя вправо, отрицательным — влево (если смотреть против направления распространения света, выходящего из николя). Тогда если знаки кристаллической пластинки $\lambda/4$ и вращения николя одинаковы, то круговая поляризация левая, если же они разные, то правая.*

3. Рассмотрим более общий вопрос: как отличить естественный свет от света, поляризованного по кругу, и от смеси естественного света с поляризованным по кругу? Для этого поставим снова на пути света пластинку в четверть волны и николь. *Если при вращении николя при любом положении пластинки интенсивность не меняется, то свет естественный. Если интенсивность меняется и падает до нуля, то он поляризован по кругу. Если же интенсивность меняется, но не падает до нуля, то падающий свет состоит из смеси естественного с поляризованным по кругу.*

4. Допустим теперь, что падающая волна *поляризована эллиптически.* Если поставить николь, то при его вращении интенсивность проходящего света в двух положениях (отличающихся друг от друга на 180°) будет максимальна, а в перпендикулярных к ним положениях минимальна. Эти положения определят направления главных осей эллипса колебаний. После этого на пути падающего света поставим пластинку $\lambda/4$, оптическая ось которой ориентирована параллельно одной из главных осей эллипса. Тогда после прохождения через пластинку свет станет поляризован линейно и может быть погашен поворотом николя. При этом николь надо будет повернуть на некоторый угол относительно исходного положения, когда интенсивность проходящего через него света была минимальна или максимальна. Действительно, в исходном положении главное сечение николя было ориентировано параллельно одной из главных осей эллипса колебаний. После же прохождения через пластинку $\lambda/4$ плоскость колебаний линейно поляризованного света будет проходить через одну из диагоналей прямоугольника на рис. 236.

5. Нетрудно сообразить, как поступить, чтобы отличить друг от друга: 1) эллиптически поляризованный свет; 2) смесь естественного света с линейно поляризованным светом (*отчасти линейно поляризованный свет*); 3) смесь естественного света с эллиптически поляризованным светом (*отчасти эллиптически поляризованный свет*). Опять надо поместить на пути распространения света пластинку в четверть волны, а за ней николь. Если вращением пластинки вокруг направления луча можно найти такое положение, при котором свет, прошедший через нее, можно погасить последующим вращением николя, то падающий свет был *эллиптически поляризован.* Если это сделать не удается, то мы имеем дело либо со смесью естественного света с линейно поляризованным, либо со смесью естественного света с эллиптически поляризованным. Чтобы отличить друг от друга эти два последних случая, на пути света ставят сначала только один николь и устанавливают его на минимум интенсивности проходящего света. Затем перед николем помещают пластинку в четверть волны. Вращением пластинки и николя снова добиваются минимума интенсивности. Если этот минимум интенсивности получается при прежнем положении николя (или при повороте его на 180°), то мы имеем *смесь естественного света с линейно поляризованным.* Если же для получения минимума требуется повернуть николь на некоторый угол, — то *смесь естественного света с эллиптически поляризованным.*

6. Вместо пластинки $\lambda/4$ (а также пластинки $\lambda/2$, вносящей дополнительную разность хода в полволны) применяются более совершенные приспособления, называемые *компенсаторами.* Они также могут служить для анализа поляризации света.

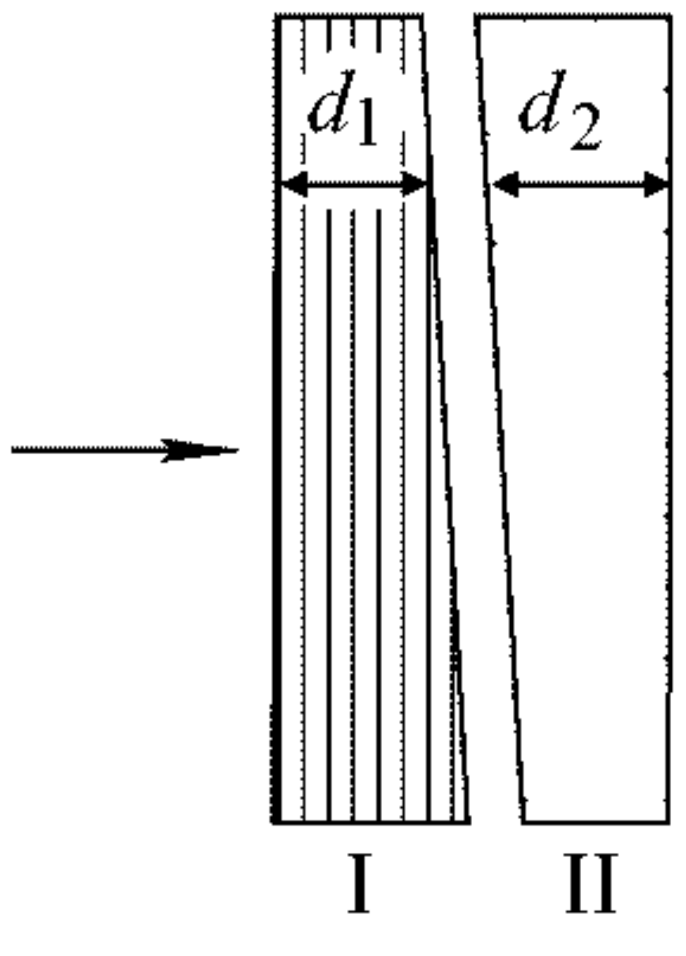


Рис. 272

Компенсатор Бабине (рис. 272) состоит из двух слабо скошенных кварцевых клиньев I и II, оптические оси которых взаимно перпендикулярны, что показано на рисунках точками и направлением штриховки. Клин I неподвижен, клин II может относительно него перемещаться вверх и вниз с помощью микрометрического винта. При таких перемещениях обращенные друг к другу поверхности клиньев остаются параллельными. Пусть луч падает на компенсатор слева направо перпендикулярно к его поверхности. В компенсаторе он разделяется на обыкновенный и необыкновенный лучи, идущие в одном направлении. Обыкновенный луч в первом клине при вступлении во второй, очевидно, становится необыкновенным и наоборот. В результате между колебаниями, параллельными и перпендикулярными к плоскости рисунка, возникнет дополнительная разность фаз

$$\Delta_{\text{доп}} = (\omega t - kn_e d_1 - kn_o d_2) - (\omega t - kn_o d_1 - kn_e d_2) = \\ = k(n_e - n_o)(d_2 - d_1),$$

где d_1 и d_2 — толщины первого и второго клиньев вдоль пересекающего их луча, а $k = 2\pi/\lambda$ — волновое число в вакууме.

Для кварца (положительный кристалл) $n_e > n_o$. В том месте, где $d_1 = d_2$, дополнительная разность фаз $\Delta_{\text{доп}}$ равна нулю. Выше этого места величина $\Delta_{\text{доп}}$ положительна, ниже — отрицательна, причем она линейно возрастает с расстоянием при смещении луча вверх. Если на компенсатор падает параллельный пучок линейно или эллиптически поляризованного света, то на выходе компенсатора в разных местах полная разность фаз Δ будет разной. Линии равной разности фаз имеют форму прямых, параллельных ребрам клиньев. Выходящий свет останется поляризованным эллиптически, но форма эллипса колебаний будет другой, меняясь при переходе от одной линии равной разности фаз к другой. Когда полная разность фаз равна $\Delta = 2m\pi$ $(m = 0, 1, 2, \ldots)$, свет будет поляризован линейно. Если его рассматривать через надлежаще ориентированный николь, то на поверхности компенсатора появятся равноотстоящие параллельные темные полосы. Если николь повернуть на 90°, то полосы сместятся на полполосы, т.е. появятся там, где $\Delta = (2m+1)\pi$. Если микрометрическим винтом перемещать клин II, то при неизменном положении луча будет меняться и дополнительная разность фаз $\Delta_{\text{доп}}$. Шкалу на барабане микрометрического винта можно проградуировать таким образом, чтобы сразу получить значение величины $\Delta_{\text{доп}}$.

Разность фаз $\Delta_{\text{доп}}$, вносимая компенсатором Бабине, зависит от точки вхождения падающего луча. В усовершенствованной форме, приданной компенсатору Солейлем (1798–1878), этот недостаток устранен: все падающие лучи приобретают одну и ту же дополнительную разность фаз. Как и компенсатор Бабине, *компенсатор Бабине–Солейля* состоит из таких же кварцевых клиньев, но ориентированных оптическими осями параллельно друг другу (рис. 273). Перед клиньями располагается плоско-параллельная кварцевая пластинка, оптическая ось которой перпендикулярна к оптической оси клиньев. Толщина пластинки равна сумме толщин клиньев в нулевом положении, когда они по высоте не смещены относительно друг друга. Оба клина вместе действуют как пло-

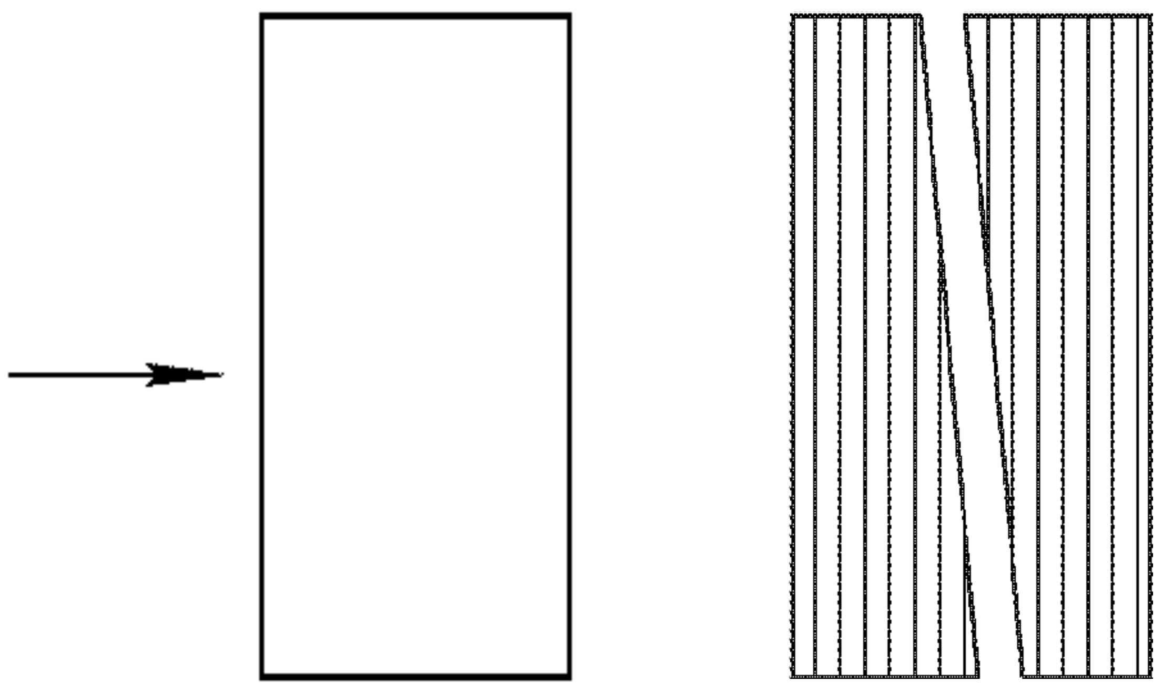


Рис. 273

скопараллельная пластинка, толщину которой можно менять микрометрическим винтом. Тем самым действие компенсатора Бабине–Солейля сведено к действию двух кристаллических пластинок с взаимно перпендикулярными оптическими осями.

ЗАДАЧИ

1. Определить наименьшую толщину d пластинки слюды, чтобы она могла служить пластинкой в четверть волны для желтого света натрия. Для такого света показатели преломления волн, идущих внутри пластинки перпендикулярно к ней, соответственно равны $n_1 = 1{,}5941$, $n_2 = 1{,}5887$. Сделать то же самое для кварца ($n_o = 1{,}5442$, $n_e = 1{,}5533$).

Ответ. 1) $d = \dfrac{\lambda}{4(n_1 - n_2)} = 27$ мкм; 2) $d = \dfrac{\lambda}{4(n_e - n_o)} = 16{,}2$ мкм.

2. Смесь света, поляризованного по кругу, и естественного рассматривается через кристаллическую пластинку в четверть волны и николь. При вращении николя вокруг оси светового пучка найдено, что максимальная интенсивность света, прошедшего через систему, в $m = 3$ раза превосходит минимальную интенсивность. Найти отношение интенсивности света $I_\text{к}$, поляризованного по кругу, к интенсивности естественного света I_e.

Ответ. $I_\text{к}/I_e = {}^1\!/_2(m-1) = 1$.

3. На пути линейно поляризованного света поставлена пластинка в полволны. Плоскость колебаний падающего света составляет угол α с оптической осью пластинки. Определить поляризацию света, прошедшего через пластинку.

Ответ. Линейная поляризация сохранится, но плоскость колебаний будет наклонена к оптической оси под углом $\beta = -\alpha$, т.е. повернется на угол 2α. В частности, когда $\alpha = 45^\circ$, плоскость колебаний повернется на 90°.

4. Параллельный пучок монохроматического света, поляризованный по правому кругу, падает нормально на пластинку в полволны. Найти состояние поляризации света, прошедшего через эту пластинку.

Ответ. Свет будет поляризован по левому кругу.

5. Параллельный пучок монохроматического света падает нормально на поляроид, а затем на пластинку в полволны. Главная плоскость поляроида (в которой лежит электрический вектор пропускаемой им волны) составляет угол α с осью этой пластинки. Найти состояние поляризации прошедшего света на выходе из пластинки в полволны.

Ответ. Свет останется линейно поляризованным, но плоскость колебаний электрического вектора повернется на угол 2α и станет симметрично расположенной со своим исходным положением относительно оси пластинки в полволны.

6. Параллельный пучок монохроматического света проходит через два николя, главные сечения которых повернуты друг относительно друга на угол $\alpha = 20^\circ$. Между николями ставится пластинка в полволны из одноосного кристалла, вырезанная параллельно оптической оси. Какой угол β должна составлять оптическая ось пластинки с главным направлением первого николя, чтобы свет через эту систему не прошел?

Решение. Свет не пройдет через второй николь, если электрический вектор перпендикулярен к главному сечению этого николя, т.е. параллелен прямой AB, перпендикулярной к тому же сечению (рис. 274). Ось пластинки должна быть ориентирована по биссектрисе угла AON_1 или угла N_1OB, ему дополнительного до π (см. предыдущую задачу). Это дает два значения угла β:

$$\beta_1 = -\frac{\pi}{4} + \frac{\alpha}{2} = -35^\circ, \quad \beta_2 = \frac{\pi}{4} + \frac{\alpha}{2} = +55^\circ.$$

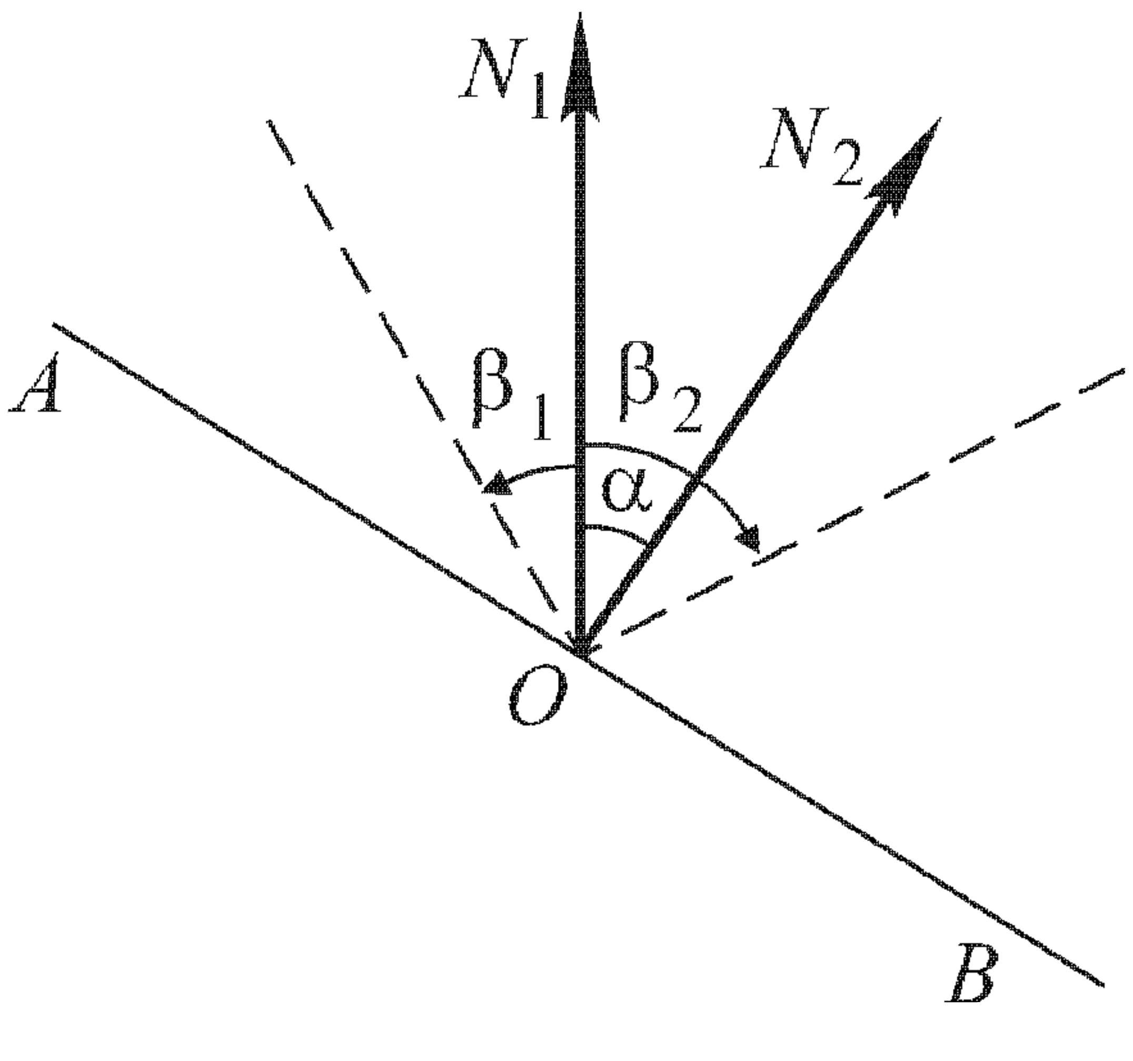


Рис. 274

7. Линейно поляризованная волна проходит через кристаллическую пластинку, одно из главных направлений которой составляет с главным сечением поляризатора угол φ. Разность фаз, сообщаемая пластинкой, равна δ. Найти: 1) отношение полуосей эллипса колебаний полученного эллиптически поляризованного света; 2) угол между главными направлениями пластинки и полуосями эллипса.

Решение. Линейно поляризованный луч по выходе из пластинки превращается в эллиптически поляризованный:

$$x = a\cos\omega t, \quad y = b\cos(\omega t + \delta),$$

где x и y — составляющие электрического вектора вдоль координатных осей, совпадающих с главными направлениями пластинки. Для определения полуосей полученного эллипса найдем проекции светового вектора x' и y' на оси координат, повернутые на угол θ относительно первоначальной системы:

$$x' = a\cos\theta\cos\omega t + b\sin\theta\cos(\omega t + \delta) = A\cos(\omega t + \alpha),$$
$$y' = -a\sin\theta\cos\omega t + b\cos\theta\cos(\omega t + \delta) = B\cos(\omega t + \beta),$$

где A и B — полуоси эллипса, определяющиеся уравнениями

$$\begin{aligned} A\cos\alpha &= a\cos\theta + b\sin\theta\cos\delta, & A\sin\alpha &= b\sin\theta\sin\delta, \\ B\cos\beta &= -a\sin\theta + b\cos\theta\cos\delta, & B\sin\beta &= b\cos\theta\sin\delta. \end{aligned} \tag{78.1}$$

Отсюда

$$\begin{aligned} A^2 &= a^2\cos^2\theta + b^2\sin^2\theta + ab\sin 2\theta\cos\delta, \\ B^2 &= a^2\sin^2\theta + b^2\cos^2\theta - ab\sin 2\theta\cos\delta. \end{aligned} \tag{78.2}$$

Складывая и вычитая эти соотношения, получим

$$A^2 + B^2 = a^2 + b^2 = R^2, \tag{78.3}$$

$$A^2 - B^2 = (a^2 - b^2)\cos 2\theta + 2ab\sin 2\theta\cos\delta, \tag{78.4}$$

где R — амплитуда волны, прошедшей через поляризатор. Чтобы амплитуды A и B соответствовали колебаниям вдоль осей эллипса, надо выбрать θ так, чтобы разность $A^2 - B^2$ была максимальной или минимальной. Приравнивая ее производную нулю, из этого условия найдем

$$\operatorname{tg} 2\theta = \frac{2ab}{a^2 - b^2}\cos\delta. \tag{78.5}$$

Из (78.1) находим

$$\begin{aligned} AB\sin(\alpha - \beta) &= -ab\sin\delta, \\ AB\cos(\alpha - \beta) &= ab\cos 2\theta\cos\delta - \frac{a^2 - b^2}{2}\sin 2\theta, \end{aligned} \tag{78.6}$$

или

$$\frac{AB\cos(\alpha-\beta)}{ab\cos 2\theta\cos\delta} = 1 - \frac{a^2-b^2}{2ab\cos\delta}\operatorname{tg}2\theta.$$

Подставляя вместо $\operatorname{tg}2\theta$ его значение из (78.5), получаем $\frac{AB\cos(\alpha-\beta)}{ab\cos 2\theta\cos\delta} = 0$, откуда $\cos(\alpha-\beta) = 0$, $(\alpha-\beta) = \pm\pi/2$. Следовательно, из первого уравнения (78.6) имеем $AB = \pm ab\sin\delta$. Так как $b/a = \operatorname{tg}\varphi$, то

$$\sin 2\varphi = \frac{2ab}{a^2+b^2},\quad \cos 2\varphi = \frac{a^2-b^2}{a^2+b^2},\quad \operatorname{tg}2\varphi = \frac{2ab}{a^2-b^2}.$$

Таким образом, уравнение (78.5) получит следующий вид:

$$\operatorname{tg}2\theta = \operatorname{tg}2\varphi\cos\delta.$$

Этим уравнением и определяется угол θ между главными направлениями пластинки и осями эллипса. Для второго неизвестного введем обозначение $B/A = \operatorname{tg}I$ и воспользуемся формулой $\sin 2I = 2\operatorname{tg}I/(1+\operatorname{tg}^2 I)$. Тогда получим

$$\sin 2I = \frac{2AB}{A^2+B^2} = \frac{2ab\sin\delta}{a^2+b^2} = \sin 2\varphi\sin\delta.$$

8. На кристаллическую пластинку, вырезанную параллельно оптической оси, падает нормально свет, поляризованный по кругу. Прошедший свет рассматривается через анализатор.

1) Пренебрегая потерями света на отражение, определить интенсивность прошедшего света, если главное сечение анализатора составляет угол α с одним из главных направлений пластинки. 2) Под каким углом надо поставить анализатор, чтобы получить максимальную и минимальную интенсивности?

Решение. 1) Если свет поляризован по кругу, то слагающие колебания по координатным осям могут быть представлены в виде

$$x = a\cos\omega t,\quad y = a\sin\omega t.$$

После прохождения через кристаллическую пластинку, сообщающую разность фаз δ, уравнения колебаний перейдут в

$$x = a\cos\omega t,\quad y = a\sin(\omega t+\delta).$$

При выходе из анализатора результирующее колебание будет

$$\xi = a\cos\alpha\cos\omega t + a\sin\alpha\sin(\omega t+\delta) = \\ = a(\cos\alpha + \sin\alpha\sin\delta)\cos\omega t + a\sin\alpha\cos\delta\sin\omega t.$$

Отсюда получаем для интенсивности

$$I = a^2\{(\cos\alpha+\sin\alpha\sin\delta)^2 + (\sin\alpha\cos\delta)^2\} = a^2(1+\sin 2\alpha\sin\delta).$$

2) При постоянном δ интенсивность достигает максимума или минимума, когда $\cos 2\alpha = 0$, т.е. при $\alpha = {}^1\!/_4\pi, {}^3\!/_4\pi$. Если $\sin\delta > 0$, то первому значению соответствует максимум, а второму — минимум; при $\sin\delta < 0$ — наоборот.

9. Клин из двоякопреломляющего вещества помещен на пути монохроматического света, поляризованного по кругу. Оптическая ось параллельна ребру клина. Описать картину, наблюдаемую через николь, когда клин неподвижен и когда он поворачивается вокруг направления распространения света.

Решение. Согласно предыдущей задаче интенсивность света, прошедшего через анализатор,

$$I = a^2(1 + \sin 2\alpha \sin\delta).$$

Если $\sin 2\alpha > 0$, то при постоянном α интенсивность минимальна, когда

$$\sin\delta = -1, \quad \text{т.е. при} \quad \delta = 3\pi/2, \quad 7\pi/2, \ldots,$$

и максимальна, когда

$$\sin\delta = 1, \quad \text{т.е. при} \quad \delta = \pi/2, \quad 5\pi/2, \quad 9\pi/2, \ldots$$

Если же $\sin 2\alpha < 0$, то в первом случае будет максимум, а во втором — минимум интенсивности. Во всех случаях в поле зрения будут видны чередующиеся светлые и темные полосы. При вращении клина будет меняться угол α, а с ним и интенсивность в каждой точке клина. При углах $\alpha = 90$, 180 и 270° весь клин будет освещен равномерно, а при углах $\alpha = 45$, 135, 225, 315° будет наблюдаться наибольший контраст темных и светлых полос. При переходе через углы $\alpha = 90$, 180, 270° темные полосы будут переходить в светлые, а светлые — в темные.

10. Два когерентных пучка квазимонохроматического неполяризованного света равной интенсивности дают на экране интерференционные полосы. Какой толщины кристаллическую пластинку надо ввести на пути одного из этих пучков, чтобы интерференционные полосы исчезли и притом так, чтобы их нельзя было восстановить никакой стеклянной пластинкой, вводимой в другой пучок? Как изменится картина, если за кристаллической пластинкой поставить поляроид? При каком положении поляроида интерференционных полос не будет?

Решение. Разложим мысленно световую волну на две составляющие, электрические векторы которых взаимно перпендикулярны и параллельны главным осям пластинки. При введении пластинки интерференционные полосы от каждой составляющей сместятся. Если введенная пластинка является пластинкой в полволны, то разность смещений составит половину ширины полосы. В этом случае при введении пластинки интерференционные полосы пропадут. При введении поляроида они появятся вновь. Исключение составляет случай, когда оси поляроида наклонены под углом 45° к осям пластинки. В этом случае интерференционные полосы наблюдаться не будут.

11. Плоская световая волна эллиптически поляризована. Длины полуосей эллипса колебаний равны соответственно a и b. Какую кристаллическую пластинку надо поставить на пути распространения волны и как надо ориентировать эту пластинку, чтобы получить свет, поляризованный по кругу: 1) с тем же направлением вращения; 2) с противоположным направлением вращения?

Решение. В системе главных осей X, Y эллиптическое колебание представляется уравнениями $E_X = a\cos\omega t$, $E_Y = b\sin\omega t$ (рис. 275). Перейдем к новой системе ξ, η, оси которой являются биссектрисами прежних координатных углов. В этой системе то же колебание представится в виде

$$E_\xi = \frac{1}{\sqrt{2}}(a\cos\omega t + b\sin\omega t) = \sqrt{\frac{a^2+b^2}{2}}\cos(\omega t - \varphi),$$

$$E_\eta = \frac{1}{\sqrt{2}}(-a\cos\omega t + b\sin\omega t) = \sqrt{\frac{a^2+b^2}{2}}\cos[\omega t - (\pi - \varphi)],$$

где φ — острый угол, определяемый уравнением $\operatorname{tg}\varphi = b/a$. Колебания вдоль осей ξ и η совершаются с одинаковыми амплитудами $\sqrt{(a^2+b^2)/2}$, причем колебание вдоль оси ξ опережает по фазе колебание вдоль оси η на угол $\delta = \pi - 2\varphi$. Внесем кристаллическую пластинку, чтобы ее оси были ориентированы вдоль ξ и η и чтобы она изменила разность фаз до $\pm\pi/2$. Для этого должно быть выполнено соотношение

$$(\omega t-\varphi-k_\xi l)-(\omega t-\pi+\varphi-k_\eta l) = \pm\pi/2,$$

откуда

$$l = \frac{2\varphi-\pi\pm\pi/2}{k_\eta-k_\xi} = \lambda\frac{\varphi/\pi-{}^1\!/_2\pm{}^1\!/_4}{n_\eta-n_\xi}.$$

Тогда волна перейдет в волну, поляризованную по кругу. Знаку плюс соответствует то же направление вращения, что и в исходной эллиптически поляризованной волне, а знаку минус — противоположное. Такой же результат получится, если толщину пластинки изменить на $m\lambda/(n_\eta - n_\xi)$, где m — целое число.

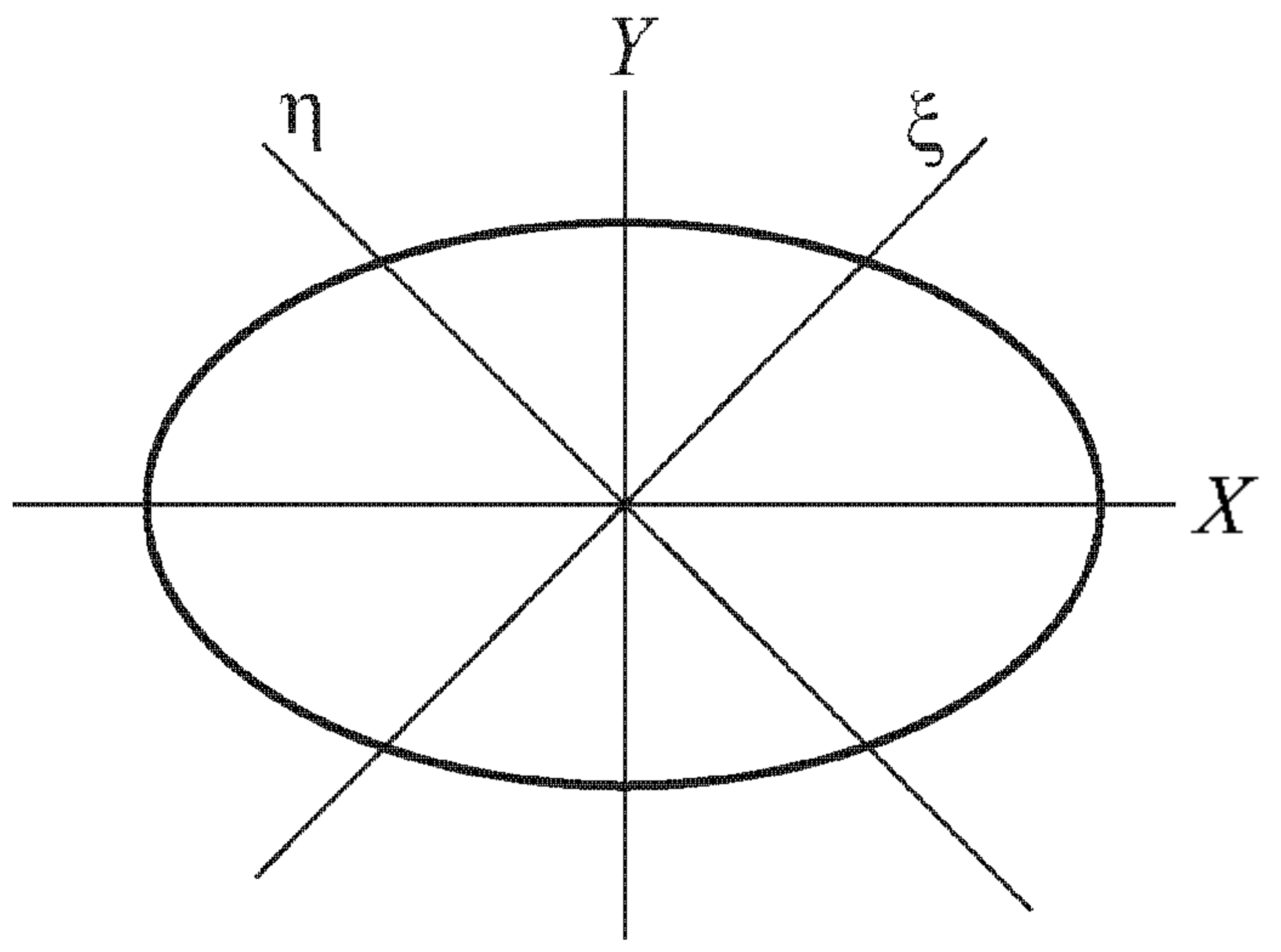


Рис. 275

§ 79. Интерференция поляризованных лучей

1. Явления *интерференции поляризованных лучей* в истории оптики имели большое значение для выяснения фундаментального вопроса о *природе световых колебаний*. Они исследовались в классических опытах Френеля и Араго (1816 г.). Конечно, лучи от независимых источников света интерферировать не будут, даже если они предварительно пропущены через поляризационное приспособление. Для интерференции необходима *когерентность*. Однако, как видно из формулы (26.2), результат интерференции линейно поляризованных лучей зависит от угла между плоскостями световых колебаний. Интерференционные полосы наиболее контрастны, когда плоскости колебаний параллельны. Интерференция никогда не наблюдается, если волны поляризованы во взаимно перпендикулярных плоскостях. Это впервые было установлено в упомянутых выше опытах Френеля и Араго. Отсюда Френель пришел к заключению о *поперечности световых колебаний* (см. § 26, п. 5).

2. Одна из возможных схем для исследования интерференции поляризованных лучей изображена на рис. 276. Лучи от первичного источника света S проходят через поляроид П или другое поляризационное приспособление. Вторичные когерентные источники S_1 и S_2 получаются одним из способов, применяемых для осуществления двухлучевой интерференции. Исходящие из них пучки поляризованы в параллельных плоскостях.

На пути одного из пучков вводится полуволновая кристаллическая пластинка K. В другом пучке для компенсации возникшей разности хода помещается стеклянная пластинка P надлежащей толщины. Стеклянная пластинка, конечно, не меняет направления колебаний проходящей через нее линейно поляризованной волны. Кристаллическая пластинка действует так же только в том случае, когда ее оптическая ось параллельна или перпендикулярна к плоскости колебаний. В этом случае из пластинок K и P выходят одинаково поляризованные когерентные пучки света. Направления колебаний в этих пучках изображены на рис. 276 совпадающими векторами $\mathbf{E}_1$ и $\mathbf{E}_2$.

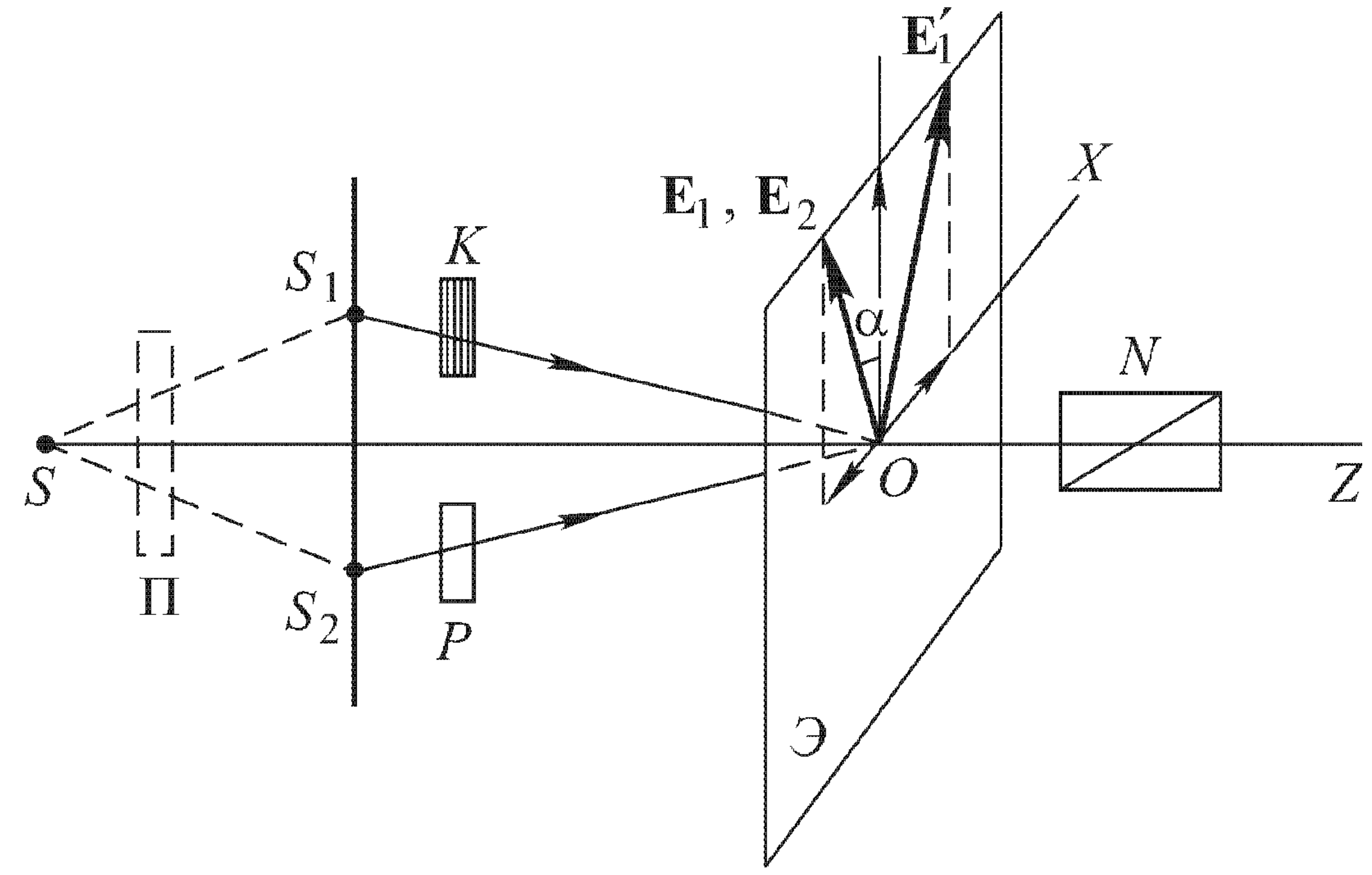


Рис. 276

Если повернуть пластинку K вокруг ее нормали на угол α, то вектор $\mathbf{E}_1$ повернется на угол 2α и перейдет в положение $\mathbf{E}_1'$ (см. задачу 5 к предыдущему параграфу). Таким образом, можно исследовать интерференцию при различных значениях угла 2α между плоскостями колебаний интерферирующих волн.

Интерференционную картину можно получать на экране Э, а также наблюдать без экрана с помощью лупы или зрительной трубы. В дальнейших пояснениях плоскость экрана принимается за координатную плоскость XY, направление оптической оси пластинки K — за ось Y, а нормаль к экрану, проведенная в сторону распространения света, — за ось Z. Следовательно, при повороте пластинки K будет поворачиваться и ось Y.

От наложения линейно поляризованных монохроматических волн везде, в том числе и на экране Э, возникнет эллиптическая поляризация света. При неизменном угле между плоскостями колебаний форма и ориентация эллипса колебаний будут изменяться с изменением разности хода между интерферирующими лучами. При определенных значениях разности хода эллиптические колебания могут вырождаться в линейные. Поверхности (а на экране — линии) равной разности хода являются в то же время поверхностями (линиями) одинаковых и одинаково ориентированных эллипсов колебаний.

Разложим, например, в описанной выше установке все колебания по осям X и Y. В центре экрана O y-составляющие векторов $\mathbf{E}_1'$ и $\mathbf{E}_2$ направлены в одну сторону, т.е. колебания вдоль оси Y совершаются в фазе. Составляющие тех же векторов вдоль оси X направлены противоположно — вдоль этой оси колебания противофазны и взаимно гасят друг друга. В результате в точке O, а также во всякой точке оси X получится линейное колебание параллельно оси Y. Если сместиться параллельно оси Y на половину ширины интерференционной полосы, то x-колебание станет синфазным, а y-колебание противофазным. Результирующее колебание также будет линейным, но уже вдоль оси X. В промежуточных точках из-за разности хода между лучами полного гашения не будет ни для x-, ни для y-колебания. Однако одно из этих колебаний всегда будет сдвинуто относительно другого по фазе на $\pi/2$: результирующее колебание — эллиптическое с главными осями эллипса, параллельными координатным осям X и Y.

3. Эллиптический характер поляризации проявляется при наблюдении через николь. Если николь пропускает колебания, параллельные оси Y, то на экране получаются интерференционные полосы, параллельные оси X. Центральная полоса, проходящая через центр O экрана (где разность хода равна нулю), будет светлой. При повороте николя на 90° эти полосы гасятся. Зато появляется другая система интерференционных полос, сдвинутая относительно предыдущей на половину ширины полосы. В ней, следовательно, центральная полоса будет темной. Таким образом, поворот николя на 90° из указанного исходного положения приводит к замене светлых полос темными и наоборот.

То же, в частности, будет, когда поляризация круговая (т.е. когда в нашей интерференционной установке кристаллическая пластинка K повернута из своего исходного положения на угол $\alpha = 45°$). Тогда интерференционные полосы, как они воспринимаются глазом, будут иметь один и тот же вид при любой ориентации николя. При вращении николя полосы непрерывно перемещаются параллельно самим себе. При повороте на 90° это перемещение составляет половину ширины полосы. Направление колебаний изменяется также на 90°. Если убрать николь, то произойдет наложение двух сдвинутых на полполосы некогерентных интерференционных картин со взаимно перпендикулярными колебаниями. Это даст равномерную освещенность экрана, в соответствии с тем, что волны со взаимно перпендикулярными колебаниями никогда не интерферируют. Таким образом, поворот плоскости колебаний одного из линейно поляризованных лучей на 90° (осуществляемый, например, вращением кристаллической пластинки K) существенно меняет распределение освещенности на экране. При отсутствии такого поворота на экране видны полосы интерференции. При наблюдении через николь их интенсивность изменяется; полосы пропадают, когда колебания перпендикулярны к главному сечению николя. При наличии же поворота плоскости колебаний одного из лучей на 90° и отсутствии николя никаких интерференционных полос не наблюдается. Зато они появляются при наблюдении через николь и непрерывно перемещаются параллельно самим себе при его вращении, причем при повороте николя на 90° это перемещение равно половине ширины полосы.

Если опыт производить в неполяризованном свете (т.е. убрать поляроид П), то введение кристаллической пластинки, повернутой на любой угол, конечно, оставляет естественным свет, проходящий через нее. Однако, как указал и подтвердил свое заключение на опыте С.И. Вавилов, при этом также должна претерпеть изменения интерференционная картина на экране. Это объясняется тем, что и после введения кристаллической пластинки лучи не остаются независимыми, а определенным образом скоррелированы между собой.

В самом деле, падающий естественный свет можно представить в виде суперпозиции линейно поляризованных некогерентных волн с всевозможными направлениями плоскости колебаний. Возьмем одну из таких волн. При наличии кристаллической пластинки K и при наблюдении через николь N на экране получатся интерференционные полосы, описанные выше. Существенно, что положение полос не зависит от угла между плоскостями колебаний интерферирующих волн. Значит, все линейно поляризованные компоненты, на которые разложена падающая волна, дадут совпадающие по положению интерференционные картины. А так как эти картины некогерентны, то их наложение приведет только к взаимному усилению интерференционных полос. При вращении николя, ввиду статистической изотропии естественного света относительно всех направлений, перпендикулярных к лучу, вид интерференционных полос изменяться не будет. Будет происходить лишь перемещение полос параллельно самим себе, как при круговой поляризации, описанной выше. При повороте николя на 90° светлые полосы заменятся темными и наоборот. Не то будет, когда нет кристаллической пластинки. В этом случае в отсутствие николя интерференционные полосы наблюдаться будут. Однако при введении николя интенсивность полос уменьшится вдвое. Вращение же николя вообще не сказывается на положении и интенсивности полос.

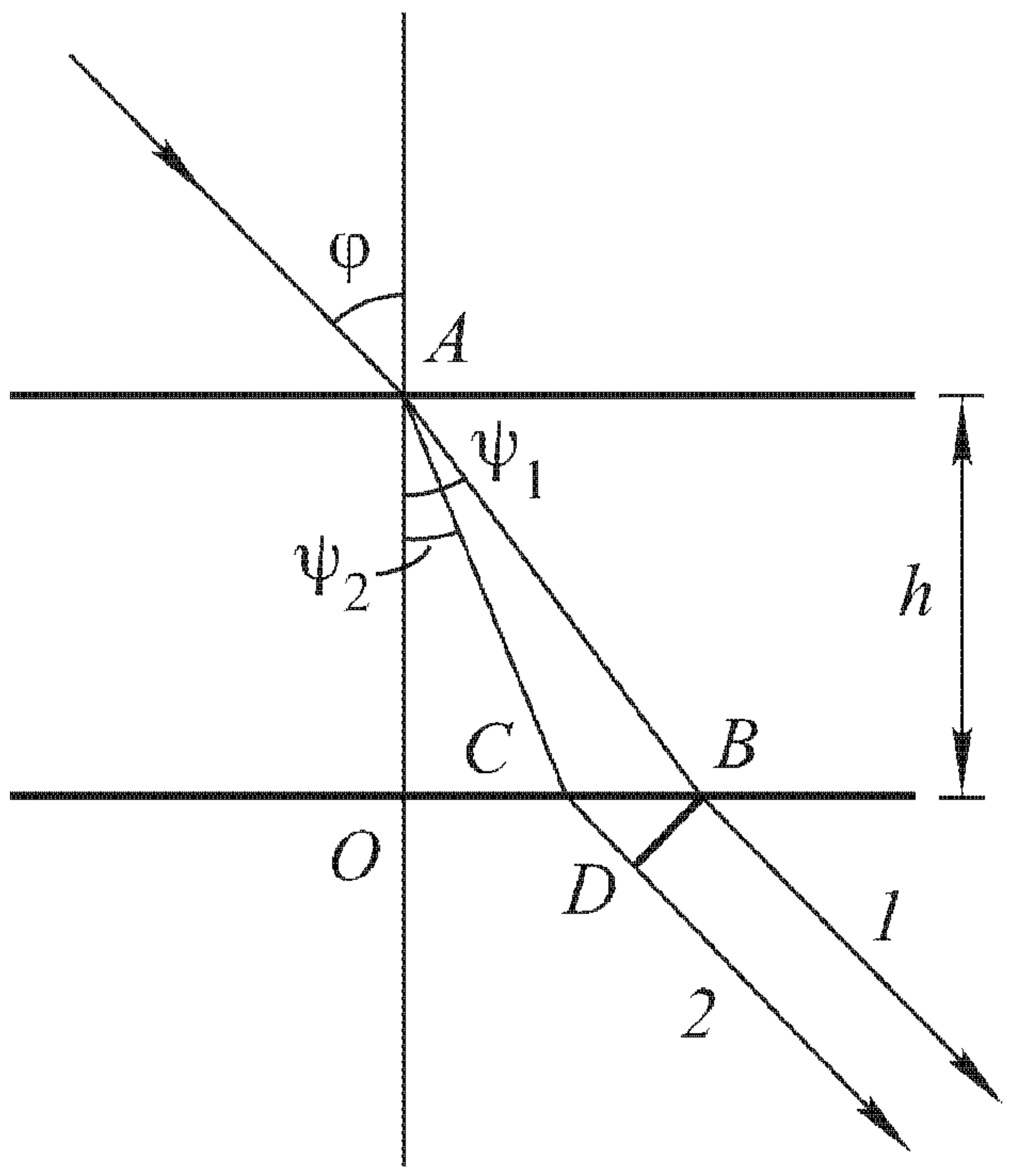


Рис. 277

4. Важный случай интерференции осуществляется при прохождении через кристалл поляризованного света. Пусть на плоскопараллельную кристаллическую пластинку падает волна под углом φ (рис. 277). В кристалле она разделяется на две волны, распространяющиеся в разных направлениях и с различными скоростями. Пусть AB и AC — волновые нормали этих волн, а ψ_1 и ψ_2 — соответствующие им углы преломления. В кристалле направления волновых нормалей не совпадают. Вне кристалла различие между этими направлениями пропадает. Из кристалла выходят два луча *1* и *2*, параллельные падающему и поляризованные в перпендикулярных плоскостях. Оптическая разность хода между ними представляется выражением

$$\Delta = (ACD) - (AB) = n_2 \cdot AC + CD - n_1 \cdot AB,$$

где n_1 и n_2 — показатели преломления рассматриваемых волн, а D — основание перпендикуляра, опущенного из точки B на луч *2*. Если h — толщина пластинки, то

$$AC = h/\cos\psi_2, \quad AB = h/\cos\psi_1, \quad CD = h(\operatorname{tg}\psi_1 - \operatorname{tg}\psi_2)\sin\varphi.$$

По закону преломления $\sin\varphi = n_1 \sin\psi_1 = n_2 \sin\psi_2$. Используя эти выражения, получим

$$\Delta = h(n_2\cos\psi_2 - n_1\cos\psi_1). \qquad (79.1)$$

Разность хода Δ между лучами *1* и *2* обусловлена двумя обстоятельствами: 1) различием показателей преломления n_1 и n_2; 2) различием углов преломления ψ_1 и ψ_2. Второе обстоятельство играет малую роль. В большинстве случаев им можно пренебречь и пользоваться приближенным выражением

$$\Delta = h\cos\psi(n_2 - n_1), \qquad (79.2)$$

где угол ψ имеет любое промежуточное значение между углами преломления ψ_1 и ψ_2.

5. Разность хода Δ можно использовать для получения интерференции лучей *1* и *2*. Если бы свет, падавший на кристаллическую пластинку, был естественный, то интерференция была бы невозможна, так как в этом случае лучи *1* и *2* не были бы коррелированы между собой. Для получения коррелированных лучей *1* и *2* падающий свет должен быть поляризован —

линейно или эллиптически. Но и в этом случае при наложении лучей *1* и *2* интерференция все же не возникнет, так как лучи *1* и *2* поляризованы во взаимно перпендикулярных плоскостях. Нужно свести колебания в этих лучах к одному направлению, т.е. предварительно пропустить их через николь. В случае плоскопараллельных пластинок лучи сводятся вместе в фокальной плоскости линзы — получаются полосы равного наклона. В случае тонких пластинок переменной толщины наблюдаются полосы равной толщины, локализованные на самих пластинках.

Схема для получения так называемой *интерференции в параллельных лучах* приведена на рис. 278. Кристаллическая пластинка K вводится между поляризатором P и анализатором A. Линейно поляризованный свет, выйдя из поляризатора P, после прохождения через кристаллическую пластинку K превращается в поляризованный эллиптически, а затем проходит через анализатор A, превращаясь снова в линейно поляризованный. Интенсивность проходящего света зависит от ориентации анализатора. При вращении последнего интенсивность I будет меняться.

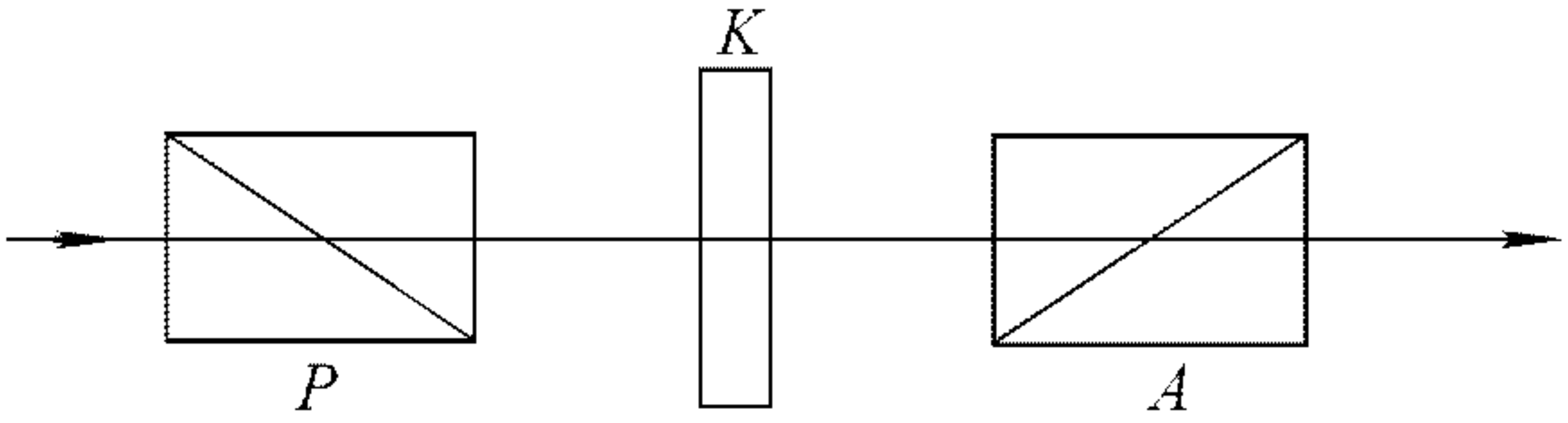


Рис. 278

Разность фаз, возникающая при прохождении света через кристаллическую пластинку, равна $2\pi\Delta/\lambda$, т.е. зависит от длины волны λ. При неподвижном анализаторе она различна для разных длин волн. По этой причине интерференционная картина получается окрашенной. При вращении анализатора окраска меняется. При повороте анализатора на 90° окраска меняется на дополнительную — светлые места становятся темными и наоборот. Если главные сечения поляризатора и пластинки параллельны или перпендикулярны между собой, то через пластинку проходит только один луч. Ему не с чем интерферировать, и окрашивание пропадает. Все эти явления получили физически не совсем удачное название *хроматической поляризации*. Она была открыта Араго в 1911 г.

Обычно главные сечения поляризатора и анализатора устанавливаются параллельно (*николи параллельны*) или перпендикулярно (*николи скрещены*) друг к другу. В последнем случае установка весьма чувствительна к обнаружению анизотропии. Через два скрещенных николя свет не проходит (поле зрения темное). Если же между ними ввести какое-либо анизотропное тело, то даже при наличии слабой анизотропии система начинает пропускать свет (поле зрения просветляется).

Для количественного расчета примем главные направления кристаллической пластинки за координатные оси X и Y (рис. 279). Пусть поляризатор пропускает колебания только в направлении P, а анализатор — только в направлении A. Углы между этими направлениями и осью X обозначим соответственно через α и β. Амплитуду света, вышедшего из поляризатора, примем за единицу. Тогда по выходе из пластинки амплитуда световых колебаний вдоль оси X будет $\cos\alpha$, а вдоль оси Y — $\sin\alpha$. Из этих колебаний анализатор пропустит колебания с амплитудами $a_1 = \cos\alpha\cos\beta$ и $a_2 = \sin\alpha\sin\beta$ соответственно, причем в результате прохождения через пластинку между обоими колебаниями возникнет разность фаз δ. От сложения обоих колебаний получится колебание, амплитуда a которого определяется соотношением

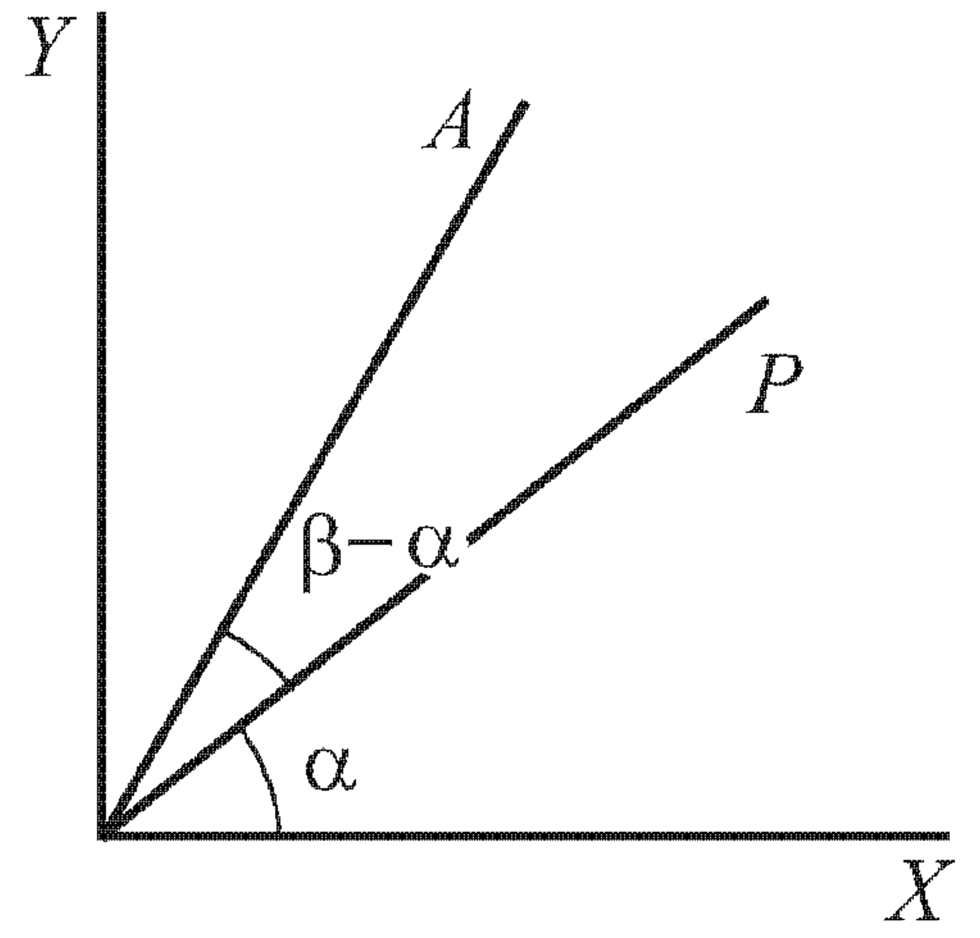


Рис. 279

$$a^2 = a_1^2 + a_2^2 + 2a_1a_2\cos\delta = (a_1+a_2)^2 - 4a_1a_2\sin^2\frac{\delta}{2}.$$

После подстановки значений a_1 и a_2, приведенных выше, для интенсивности света, проходящего через анализатор, получим

$$I = \cos^2(\beta-\alpha) - \sin 2\alpha\sin 2\beta\sin^2\frac{\delta}{2}. \tag{79.3}$$

В этой формуле в сочетании с формулой (79.1) содержится объяснение всех особенностей интерференции поляризованного света в параллельных лучах. Пока свет монохроматичен, а толщина пластинки всюду одинакова, все величины в (79.3) постоянны, так что получается *равномерная освещенность* пластинки. При повороте поляризатора или анализатора на 90° выражение (79.3) переходит в

$$I' = \sin^2(\beta-\alpha) + \sin 2\alpha\sin 2\beta\sin^2\frac{\delta}{2}.$$

Следовательно, $I + I' = 1$. Это очевидно и из общих соображений, так как при указанном повороте весь свет, пропускавшийся ранее, будет задержан, а задержавшийся начнет проходить. В частном случае, когда николи параллельны ($\alpha = \beta$),

$$I = 1 - \sin^2 2\alpha\sin^2\frac{\delta}{2}, \tag{79.4}$$

а когда скрещены —

$$I' = \sin^2 2\alpha\sin^2\frac{\delta}{2}. \tag{79.5}$$

В белом свете, если его разложить на спектральные составляющие, первое слагаемое в (79.3) остается постоянным, а второе зависит от длины волны и приводит к окрашиванию поля зрения. Если направление P совпадает с одним из главных направлений пластинки ($\alpha = 0$ или $\alpha = \pi/2$), то $I = \cos^2(\beta - \alpha)$, т.е. интенсивность I постоянна и не зависит от длины волны. В этом случае, как было уже отмечено выше, интерференции нет, и в белом свете получается равномерно освещенное неокрашенное поле. Интенсивность его можно менять поворотом николя-анализатора A, но оно все время остается белым.

6. Более сложные явления наблюдаются при *интерференции сходящихся поляризованных лучей.* Обычно применяемая схема для наблюдения таких явлений представлена на рис. 280. Параллельный пучок линейно поляризованного света линзой L_1 превращается в сходящийся и концентрируется на кристаллической пластинке K. После прохождения через пластинку K каждый луч разделяется на два параллельных луча, между которыми возникает разность хода. При наложении таких лучей свет становится поляризованным эллиптически, причем разность хода между лучами, а с ней и характер эллиптической поляризации зависят от угла наклона падающего луча. Линза L_2, одинаковая с L_1, делает лучи снова параллельными. Распределение светового поля в фокальной плоскости FF линзы проецируется объективом O и николем-анализатором A на экран, где и наблюдается интерференционная картина.

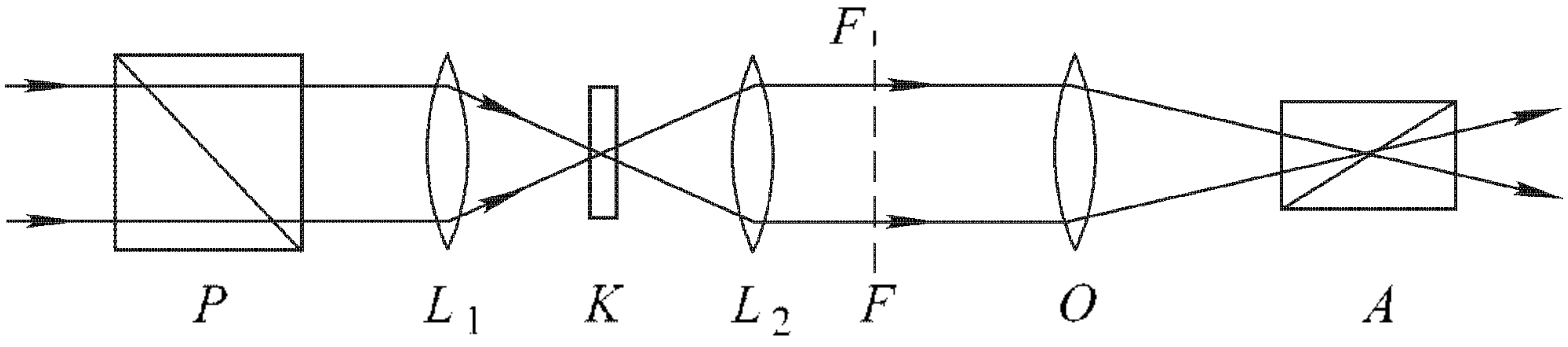


Рис. 280

Чтобы составить представление о форме получающихся интерференционных полос, введем понятие *изохроматической поверхности* (поверхности одинакового цвета). Возьмем в неограниченной кристаллической среде произвольную точку O и представим, что через нее во всевозможных направлениях и притом в одинаковых фазах проходят плоские волны. В каждом направлении волновой нормали распространяются две волны, линейно поляризованные во взаимно перпендикулярных плоскостях. На расстоянии r от точки O между ними возникнет разность хода $\Delta = (n_2 - n_1)r$, где n_1 и n_2 — показатели преломления этих волн. Фиксировав Δ, будем откладывать на волновых нормалях отрезки с длиной $r = \Delta/(n_2 - n_1)$. Геометрическое место концов таких отрезков есть поверхность равной разности фаз между волнами, которые могут распространяться вдоль одной и той же волновой нормали. Она и называется *изохроматической поверхностью.* Придавая Δ всевозможные значения, получим *семейство изохроматических поверхностей* с общим центром O. Линии, получающиеся от сечения такого семейства плоскостями, называются *изохроматическими линиями*, или *изохроматами.* Очевидно, изохроматы есть *линии постоянной разности фаз* между соответствующими волнами, могущими распространяться вдоль одной и той же волновой нормали.

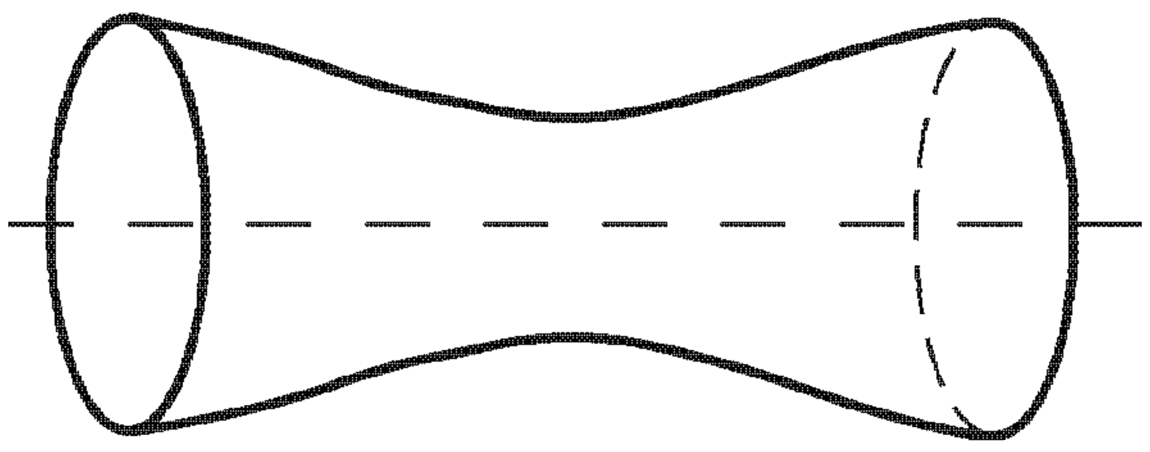


Рис. 281

Для одноосного кристалла изохроматическая поверхность есть поверхность вращения вокруг оптической оси (рис. 281). В направлении оптической оси она уходит в бесконечность, так как для этого направления оба показателя преломления n_1 и n_2 совпадают. В двуосных кристаллах таких направлений два (рис. 282). Это есть оптические оси, точнее — оси нормалей (см. § 80). Мы не будем заниматься выводом уравнения изохроматической поверхности, а ограничимся качественными соображениями.

Рис. 282

Напомним, что интерферирующие волны возникают при двойном преломлении одной и той же падающей волны. Волновые нормали получившихся двух волн внутри кристаллической пластинки несколько отличаются друг от друга по направлению. Однако таким различием мы пренебрежем, как это уже делалось при замене точной формулы (79.1) приближенной (79.2). В этом приближении, как показывается ниже, интерференционные полосы можно отождествить с определенным семейством изохромат. В самом деле, представим, что на передней плоскости кристаллической пластинки K (см. рис. 280) помещена маленькая диафрагма. (В такой диафрагме нет надобности, если задний фокус линзы L_1 находится на передней поверхности пластинки.) Примем центр этой диафрагмы за центр O семейства изохроматических поверхностей. Задняя плоскость пластинки K пересечет это семейство вдоль изохроматических линий. Последние в указанном выше приближении будут совпадать с интерференционными линиями на задней поверхности пластинки K. Если оптическая система дает изображение этой поверхности на экране, то при этом получится и подобное изображение рассматриваемой системы интерференционных линий. При иной фокусировке, из-за преломления волновых нормалей при выходе из пластинки K, полного подобия между изохроматами и интерференционными линиями на экране не будет. Однако это обстоятельство не сказывается существенно на виде интерференционных полос. Изохроматы практически всегда качественно верно передают форму интерференционных полос.

Форма изохромат и интерференционных полос определяется симметрией кристалла и ориентацией его оптических осей. На рис. 283 показаны интерференционные полосы в случае пластинки одноосного кристалла, вырезанной перпендикулярно к оптической оси. В соответствии с осевой симметрией они имеют форму концентрических кругов. Картина получена в скрещенных николях. Кольца пересечены темным крестом. Происхождение его весьма простое. На пластинку K падает линейно поляризованный свет. Рассмотрим плоскость главного сечения и плоскость, перпендикулярную к ней. Пусть обе плоскости пересекаются вдоль оси вращения системы. Лучи в этих плоскостях не испытывают двойного преломления, интерференции в таких лучах не возникает. После прохождения через пластинку обе системы лучей остаются линейно поляризованными с сохранением прежнего направления колебаний. Лучи с такими колебаниями гасятся николем-анализатором. Вокруг прямых, по которым рассматриваемые плоскости пересекаются с плоскостью экрана, и формируется темный крест. Если николи перевести в параллельное положение, то вся картина сменяется дополнительной, в частности, черный крест заменяется белым.

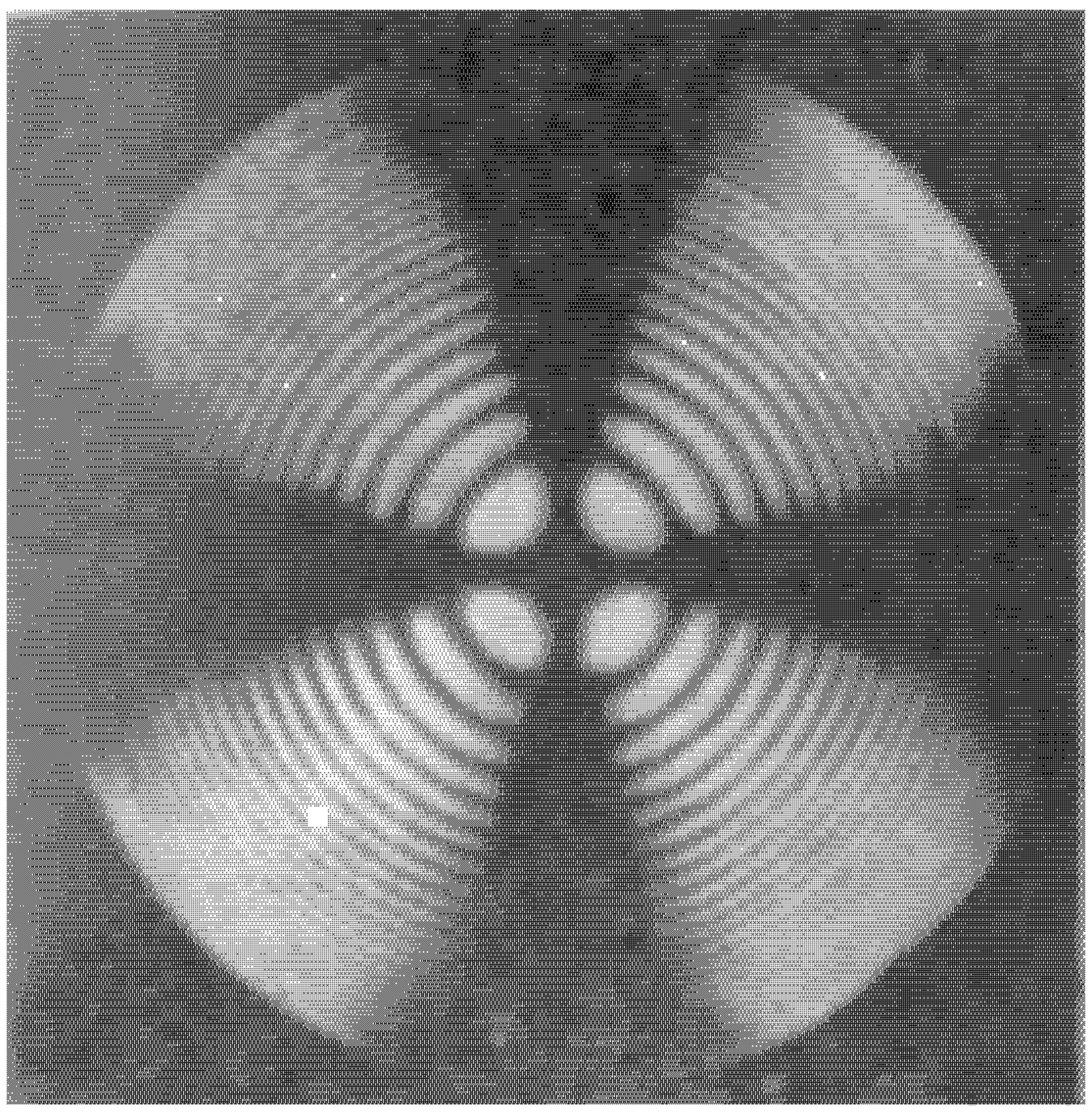


Рис. 283

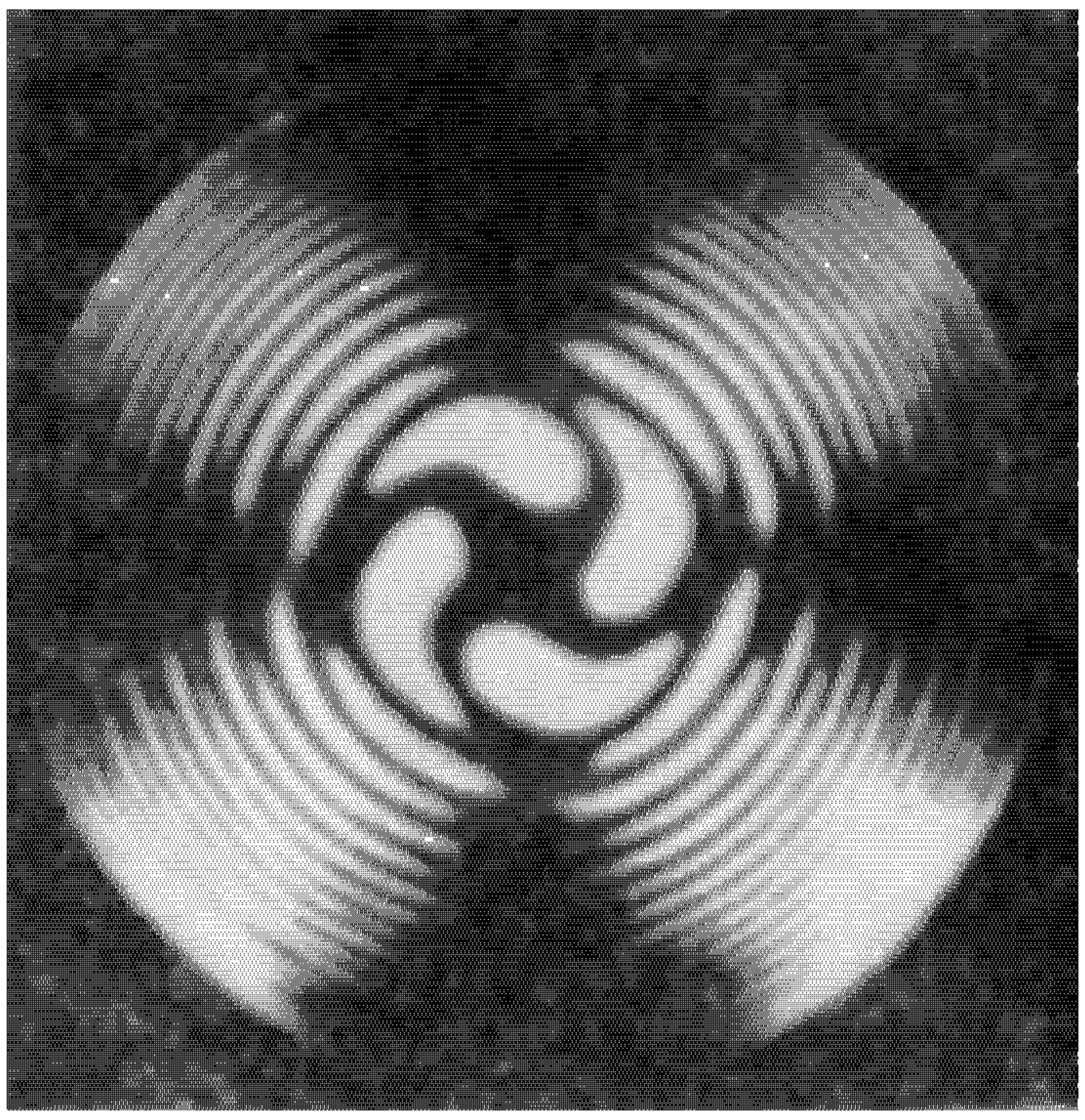


Рис. 284

Если в предыдущем опыте пластинку K заменить двумя сложенными вместе пластинками из правого и левого кварца, вырезанными перпендикулярно к оптической оси, то получаются спиралеобразные интерференционные картины, называемые *спиралями Эйри* (рис. 284). Их происхождение связано с тем, что кварц *вращает плоскость поляризации*. Для правого кварца это вращение происходит в одну сторону, а для левого — в противоположную.

На рис. 285 показан вид изохромат для пластинки одноосного кристалла, вырезанной параллельно оптической оси, а на рис. 286 — для пластинки двуосного кристалла, вырезанной перпендикулярно к биссектрисе угла между оптическими осями.

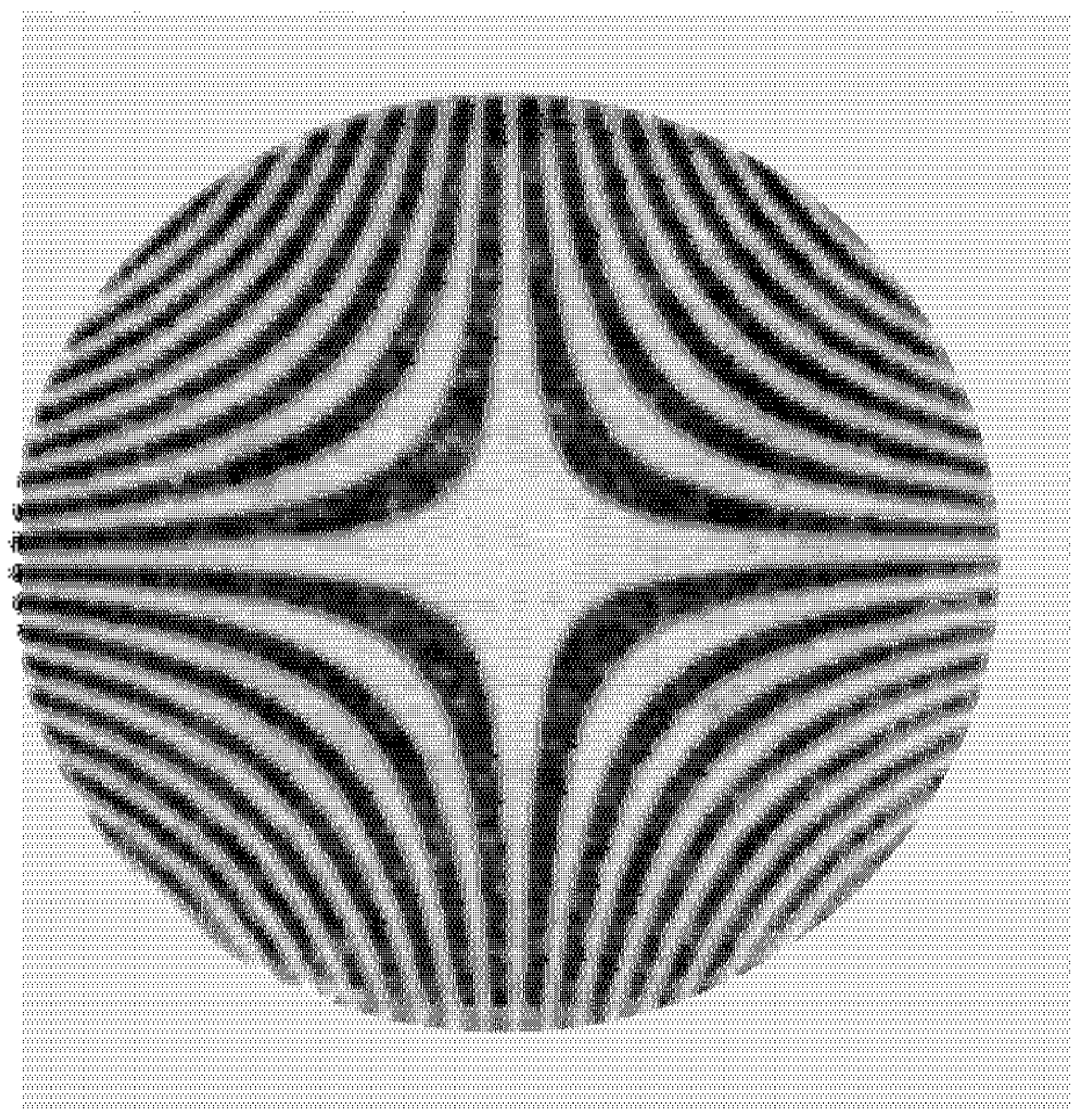


Рис. 285

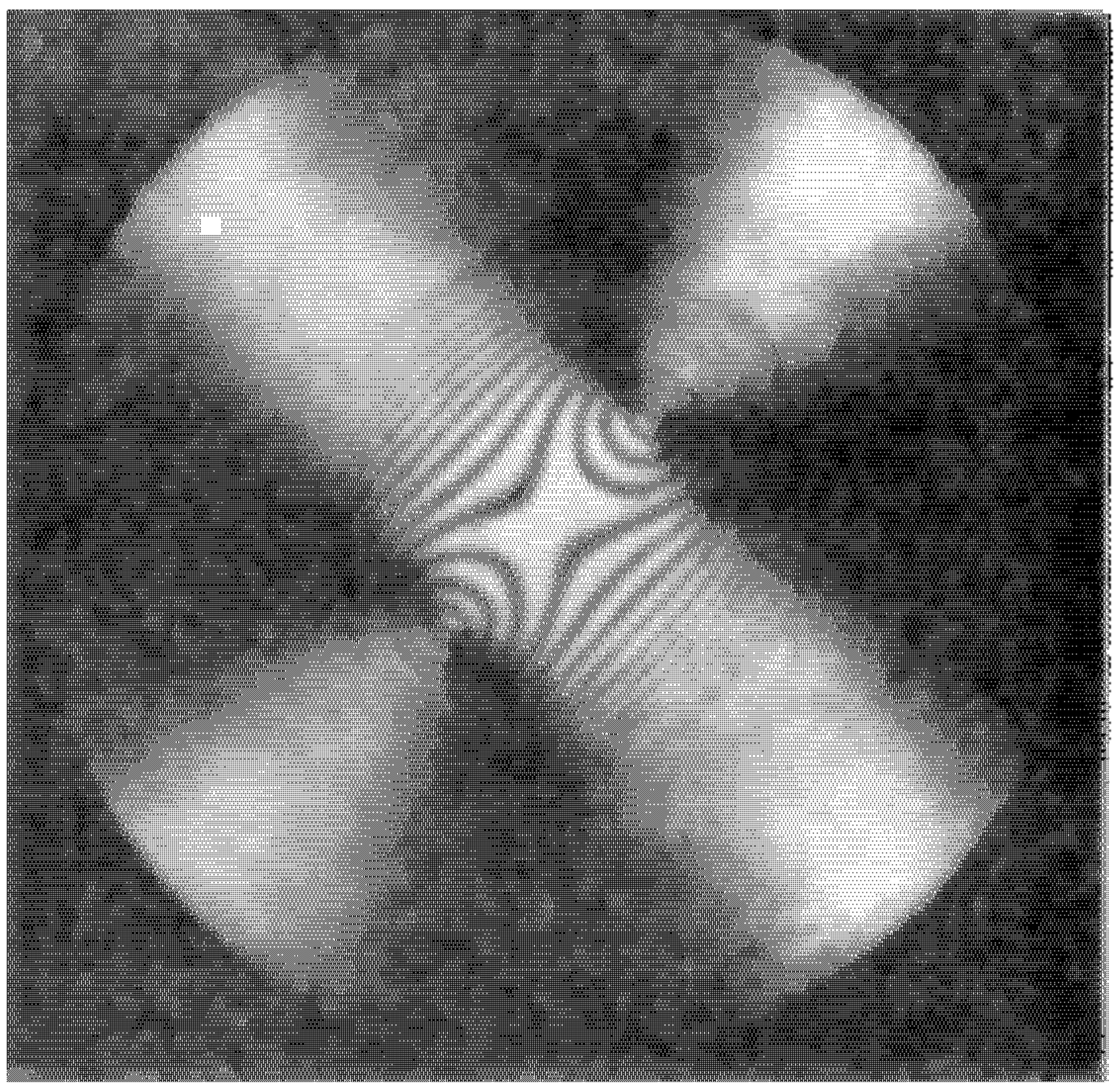


Рис. 286

7. Интерференция поляризованных лучей дает удобный и чувствительный метод обнаружения и исследования оптической анизотропии тел во всех случаях, когда она мала. Для этого обычно применяется установка в параллельных лучах с скрещенными николями, схематически изображенная на рис. 278. Вместо кристаллической пластинки ставится исследуемый образец. В отсутствие образца свет через систему не проходит. При введении образца, обладающего анизотропией, поле зрения просветляется и в случае белого света становится окрашенным. Таким методом исследуются различные случаи *искусственной анизотропии.*

Брюстером в 1815 г. было открыто *двойное преломление при механических деформациях* изотропных тел. Среда может состоять из анизотропных молекул. В отсутствие деформаций они ориентированы хаотически, вследствие чего среда макроскопически изотропна. При деформациях может возникнуть преимущественная ориентация и изменение расположения молекул в пространстве. Это и ведет к возникновению двойного преломления. В простейшем случае одностороннего растяжения или сжатия тело ведет себя как одноосный кристалл с оптической осью вдоль растяжения или сжатия. Как показали опыты, разность $n_o - n_e$ между обыкновенным и необыкновенным показателями преломления пропорциональна приложенному напряжению, а коэффициент пропорциональности зависит от длины волны. Эффект сохраняется и при остаточных деформациях. Он применяется для исследования упругих напряжений и остаточных деформаций в телах. Двойное преломление возникает при течении жидкостей с анизотропными молекулами (эффект Максвелла), а также при внесении тел в электрическое или магнитное поля (см. § 90 и 91).

ЗАДАЧИ

1. Почему никогда не может получиться интерференция обыкновенного и необыкновенного лучей, вышедших из пластинки двоякопреломляющего кристалла, настолько тонкой, что она не дает заметного разделения лучей?

2. Почему демонстрационные опыты по интерференции поляризованных лучей удобнее делать с тонкими, а не с толстыми пластинками? Почему даже с тонкими пластинками из исландского шпата трудно получить интерференционную картину в белом свете?

Ответ. У исландского шпата очень велика разность обыкновенного и необыкновенного показателей преломления, так что даже в тонких пластинках получаются большие разности хода, при которых интерференция в белом свете невозможна.

3. Две толстые пластинки одноосного кристалла, одинаково ориентированные и весьма мало отличающиеся по толщине, в скрещенных николях дают порознь белый свет. Почему в тех же условиях может получиться окрашивание, если повернуть одну пластинку относительно другой на 90°?

Решение. Когда пластинки одинаково ориентированы, вносимая ими разность хода велика, т.е. соответствует высокому порядку интерференции. В таких условиях интерференция в белом свете наблюдаться не может. Если одну из пластинок повернуть относительно другой на 90°, то вносимая ими разность хода будет такая же, какая вносится пластинкой с толщиной, равной разности толщин рассматриваемых пластинок. При малой разности толщин разность хода может сделаться настолько малой, что станет возможна интерференция в белом свете.

4. Кварцевая пластинка толщиной в 1 мм вырезана перпендикулярно к оптической оси. Как определить, из право- или левовращающего кварца сделана пластинка, имея в своем распоряжении два николя и источник: 1) монохроматического света; 2) белого света?

Ответ. 1) Если поместить пластинку кварца, вырезанную перпендикулярно к оптической оси, между скрещенными николями и осветить систему монохроматическим светом, то она будет пропускать свет; повернув анализатор на угол, меньший 90°, можно снова погасить свет. Если при этом наблюдатель должен вращать анализатор по направлению часовой стрелки, то кварц будет правовращающий, если же против часовой стрелки, то левовращающий.

2) Если осветить систему белым светом, то пластинка будет казаться окрашенной. Вращательная способность увеличивается с уменьшением

длины волны. Поэтому если вращать анализатор по часовой стрелке, то для правовращающего кварца окраска будет меняться в сторону коротких длин волн спектра. Для левовращающего кварца порядок изменения окраски будет обратный.

§ 80. Нормальные скорости и поляризация волн в двуосных кристаллах

1. Перейдем теперь к исследованию распространения волн в *оптически двуосных кристаллах.* В общем случае вектор **D** может зависеть не только от вектора **E**, но и от его пространственных производных. Это явление называется *пространственной дисперсией* (см. § 96). В слабых полях такая зависимость, конечно, может считаться линейной. Для плоских монохроматических волн дифференцирование **E** по координатам x, y, z сводится к умножению его проекций на $-ik_x$, $-ik_y$, $-ik_z$. В этом случае зависимость от пространственных производных можно учесть прежней формулой (75.2), если диэлектрический тензор ε_{kj} считать комплексным. Формально так можно поступать и в случае неплоских волн. Однако волны должны предполагаться монохроматическими.

Для непоглощающих сред диэлектрический тензор должен быть *эрмитовым*, т.е. $\varepsilon_{kj} = \varepsilon^*_{jk}$. Действительно, для производной плотности электромагнитной энергии u по времени электродинамика дает

$$\frac{\partial u}{\partial t} = \frac{1}{4\pi}(\mathbf{E}\dot{\mathbf{D}} + \mathbf{H}\dot{\mathbf{B}})$$

(см. т. III, § 84). В случае монохроматического поля в непоглощающей среде среднее значение этой производной, согласно закону сохранения энергии, должно равняться нулю. Если пользоваться комплексной формой монохроматического поля, то это условие запишется в виде $(\mathbf{E}\dot{\mathbf{D}}^* + \mathbf{H}\dot{\mathbf{B}}^*)$ + компл. сопр. = 0. А так как мы пренебрегаем различием между **B** и **H**, то $\mathbf{E}\dot{\mathbf{D}}^*$ + + компл. сопр. = 0. Отсюда с учетом соотношений $\dot{\mathbf{D}} = i\omega\mathbf{D}$, $\dot{\mathbf{D}}^* = -i\omega\mathbf{D}^*$ получаем: $\mathbf{E}\mathbf{D}^* - \mathbf{E}^*\mathbf{D} = 0$, или

$$\sum E_\alpha \varepsilon^*_{\alpha\beta} E^*_\beta - \sum E^*_\alpha \varepsilon_{\alpha\beta} E_\beta = 0.$$

Заменим в первой сумме немой индекс α на β и наоборот. Тогда

$$\sum (\varepsilon_{\alpha\beta} - \varepsilon^*_{\beta\alpha}) E^*_\alpha E_\beta = 0.$$

Это соотношение должно выполняться для любого поля **E**, что возможно тогда и только тогда, когда $\varepsilon_{\alpha\beta} = \varepsilon^*_{\beta\alpha}$. Действительно, пусть все компоненты вектора **E**, за исключением одной E_α, равны нулю. Тогда предыдущее соотношение переходит в $(\varepsilon_{\alpha\alpha} - \varepsilon^*_{\alpha\alpha})E_\alpha E^*_\alpha = 0$, откуда $\varepsilon_{\alpha\alpha} = \varepsilon^*_{\alpha\alpha}$. Пусть теперь отличны от нуля две компоненты E_α и E_β, а третья компонента равна нулю. Тогда

$$(\varepsilon_{\alpha\beta} - \varepsilon^*_{\beta\alpha})E^*_\alpha E_\beta + (\varepsilon_{\beta\alpha} - \varepsilon^*_{\alpha\beta})E_\alpha E^*_\beta = 0.$$

Полагая здесь $E_\alpha = E_\beta$, получим

$$(\varepsilon_{\alpha\beta} - \varepsilon^*_{\beta\alpha}) + (\varepsilon_{\beta\alpha} - \varepsilon^*_{\alpha\beta}) = 0.$$

Полагая же $E_\beta = iE_\alpha$, найдем

$$(\varepsilon_{\alpha\beta} - \varepsilon^*_{\beta\alpha}) - (\varepsilon_{\beta\alpha} - \varepsilon^*_{\alpha\beta}) = 0.$$

Из этого и предыдущего соотношений следует: $\varepsilon_{\alpha\beta} = \varepsilon^*_{\beta\alpha}$. Таким образом, соотношение $\varepsilon_{\alpha\beta} = \varepsilon^*_{\beta\alpha}$ справедливо как для одинаковых, так и для разных индексов α и β, т.е. для непоглощающих кристаллов тензор $\varepsilon_{\alpha\beta}$ эрмитов. Для поглощающих кристаллов он не эрмитов.

Допустим теперь, что среда не обладает пространственной дисперсией или этим явлением, ввиду его малости, можно пренебречь. Тогда величины $\varepsilon_{\alpha\beta}$ вещественны, а потому $\varepsilon_{\alpha\beta} = \varepsilon_{\beta\alpha}$, т.е. тензор $\varepsilon_{\alpha\beta}$ будет *симметричен*. В дальнейшем мы ограничимся этим случаем. Кроме того, будем предполагать, что все диагональные элементы $\varepsilon_{\alpha\alpha}$ *положительны*. Только тогда среда будет *прозрачной*, т.е. плоские волны в ней будут распространяться без затухания. В противном случае возникнет затухание без поглощения, как это имеет место, например, в плазме (см. § 87).

Всякий симметричный тензор можно привести к так называемому *диагональному виду*, т.е. найти такую систему прямоугольных координат, в которой недиагональные компоненты тензора обращаются в нуль. Диагональные компоненты тензора в этой системе координат условимся обозначать через ε_x, ε_y, ε_z, т.е. характеризовать их не двойными, а только единичными индексами x, y, z. В рассматриваемой системе материальные уравнения имеют вид

$$D_\alpha = \varepsilon_\alpha E_\alpha. \tag{80.1}$$

Координатные оси, относительно которых тензор $\varepsilon_{\alpha\beta}$ диагонален, называются *главными осями тензора* или *диэлектрическими осями кристалла*, а величины ε_x, ε_y, ε_z — *главными диэлектрическими проницаемостями*. Эти оси мы и примем за координатные оси, причем названия осей X, Y, Z установим так, чтобы соблюдались неравенства

$$\varepsilon_x \leqslant \varepsilon_y \leqslant \varepsilon_z. \tag{80.2}$$

Так как компоненты тензора $\varepsilon_{\alpha\beta}$ могут зависеть от длины волны λ, могут зависеть от λ и направления диэлектрических осей. Это явление, называемое *дисперсией диэлектрических осей*, действительно встречается в триклинных и моноклинных кристаллах, характеризующихся наиболее низкой симметрией.

Отметим, наконец, что угол α между векторами $\mathbf{E}$ и $\mathbf{D}$ всегда острый. Это вытекает из того, что скалярное произведение $(\mathbf{ED}) = ED\cos\alpha$ пропорционально плотности электрической энергии, а она существенно положительна.

2. Обратимся теперь к исследованию плоских волн (75.3) в прозрачных кристаллах в общем виде. Фиксируем направление волновой нормали $\mathbf{N}$ и определим, какие плоские волны могут распространяться в этом направлении. Используя материальные уравнения (80.1), перепишем соотношение (75.7) в следующем виде:

$$(v^2 - a_\alpha^2)D_\alpha = -c^2(\mathbf{NE})N_\alpha, \tag{80.3}$$

где введено обозначение

$$a_\alpha = c/\sqrt{\varepsilon_\alpha}. \tag{80.4}$$

Разделив на $v^2 - a_\alpha^2$, получим

$$D_\alpha = -\frac{c^2}{v^2 - a_\alpha^2}(\mathbf{NE})N_\alpha. \tag{80.5}$$

Умножим обе части этого соотношения на N_α и просуммируем по α. Тогда

$$\sum_\alpha D_\alpha N_\alpha \equiv (\mathbf{DN}) = -c^2(\mathbf{NE})\sum_\alpha \frac{N_\alpha^2}{v^2 - a_\alpha^2} = 0,$$

так как $(\mathbf{DN}) = 0$. Скалярное произведение $(\mathbf{NE})$, вообще говоря, отлично от нуля. Поэтому

$$\sum \frac{N_\alpha^2}{v^2 - a_\alpha^2} = 0. \tag{80.6}$$

В развернутом виде

$$\frac{N_x^2}{v^2 - a_x^2} + \frac{N_y^2}{v^2 - a_y^2} + \frac{N_z^2}{v^2 - a_z^2} = 0, \tag{80.7}$$

а после освобождения от знаменателей

$$F(v^2) \equiv N_x^2(v^2 - a_y^2)(v^2 - a_z^2) + N_y^2(v^2 - a_z^2)(v^2 - a_x^2) + \\ + N_z^2(v^2 - a_x^2)(v^2 - a_y^2) = 0. \tag{80.8}$$

Хотя уравнение (80.8) мы и получили преобразованием уравнения (80.7), в действительности оно обладает большей общностью. Как видно из вывода, при получении (80.7) надо было вводить предположение, что $v^2 - a_\alpha^2 \neq 0$ и $(\mathbf{NE}) \neq 0$. Уравнение (80.7) теряет смысл, когда по крайней мере одна из разностей

$v^2 - a_\alpha^2$ обращается в нуль Уравнение же (80.8) остается справедливым и в этом случае, как показывает несложное математическое исследование, которое мы опускаем.

3. Уравнение (80.6), или (80.8), называется *законом Френеля для нормальной скорости распространения световых волн в кристалле*. Если задать направление $\mathbf{N}$, то из этих уравнений можно определить нормальную скорость v. Уравнение (80.8) второй степени относительно v^2. Докажем, что оно имеет вещественные и притом положительные корни. Для прозрачных кристаллов главные диэлектрические проницаемости, а с ними и величины a_α^2, существенно положительны. При этом ввиду условия (80.2)

$$a_x \geqslant a_y \geqslant a_z. \tag{80.9}$$

Придавая в функции $F(v^2)$ аргументу v^2 значения a_x^2, a_y^2, a_z^2, придем к неравенствам:

$$\begin{aligned} F(a_x^2) &= N_x^2(a_x^2 - a_y^2)(a_x^2 - a_z^2) \geqslant 0, \\ F(a_y^2) &= N_y^2(a_y^2 - a_z^2)(a_y^2 - a_x^2) \leqslant 0, \\ F(a_z^2) &= N_z^2(a_z^2 - a_x^2)(a_z^2 - a_y^2) \geqslant 0. \end{aligned}$$

Из них видно, что функция $F(v^2)$ дважды меняет знак: один раз между a_x^2 и a_y^2, другой — между a_y^2 и a_z^2. Следовательно, уравнение $F(v^2) = 0$ имеет два вещественных положительных корня: v_1^2 и v_2^2, причем

$$a_x \geqslant v_1 \geqslant a_y \geqslant v_2 \geqslant a_z. \tag{80.10}$$

Отсюда следует, что в направлении $\mathbf{N}$ могут распространяться две волны: одна с нормальной скоростью v_1, а другая с v_2. В частных случаях скорости v_1 и v_2 могут совпадать.

Если скорости v_1 и v_2 различны, то каждая из волн будет поляризована линейно. Это следует из соотношения

$$D_x : D_y : D_z = \frac{N_x}{v^2 - a_x^2} : \frac{N_y}{v^2 - a_y^2} : \frac{N_z}{v^2 - a_z^2}, \tag{80.11}$$

которое получается из (80.5) и в котором под v следует понимать либо v_1, либо v_2. Все величины, стоящие в правой части (80.11), вещественны. Значит, между компонентами D_x, D_y, D_z нет сдвигов фаз, отличающихся от 0 или π, а потому волна поляризована линейно.

Докажем, что если скорости v_1 и v_2 различны, то векторы $\mathbf{D}$ обеих волн, которые могут распространяться в направлении $\mathbf{N}$, взаимно перпендикулярны. Отмечая величины, относящиеся к одной из волн, индексом 1, а к другой — индексом 2, из (75.7) получим

$$v_1^2\mathbf{D}_1 - c^2\mathbf{E}_2 = -c^2(\mathbf{N}\mathbf{E}_1)\mathbf{N},$$
$$v_2^2\mathbf{D}_2 - c^2\mathbf{E}_2 = -c^2(\mathbf{N}\mathbf{E}_2)\mathbf{N}.$$

Умножим первое уравнение скалярно на $\mathbf{D}_2$, второе на $\mathbf{D}_1$ и вычтем одно уравнение из другого. Так как $(\mathbf{D}_1\mathbf{N}) = (\mathbf{D}_2\mathbf{N}) = 0$, то в результате получим

$$(v_1^2 - v_2^2)\mathbf{D}_1\mathbf{D}_2 = c^2(\mathbf{E}_1\mathbf{D}_2 - \mathbf{E}_2\mathbf{D}_1).$$

Но, очевидно, $\mathbf{E}_1\mathbf{D}_2 = \mathbf{E}_2\mathbf{D}_1$, так как каждое из этих скалярных произведений равно $\sum \varepsilon_\alpha E_{1\alpha}E_{2\alpha}$. Следовательно,

$$(v_1^2 - v_2^2)\mathbf{D}_1\mathbf{D}_2 = 0.$$

Отсюда при $v_1 \neq v_2$ следует $(\mathbf{D}_1\mathbf{D}_2) = 0$, что и требовалось доказать. Аналогично докажем, что $(\mathbf{B}_1\mathbf{B}_2) = 0$.

Итак, в *каждом направлении в кристалле могут распространяться две линейно поляризованные волны, скорости которых, вообще говоря, различны. Обе волны поперечны относительно векторов* $\mathbf{D}$ *и* $\mathbf{B}$. *Векторы* $\mathbf{D}$ (*а также* $\mathbf{B}$) *в этих волнах взаимно перпендикулярны. Относительно вектора* $\mathbf{E}$ *обе волны в кристалле не поперечны, за исключением тех случаев, когда вектор* $\mathbf{E}$ *параллелен одной из диэлектрических осей кристалла.* Однако деление волн на обыкновенную и необыкновенную возможно *только для одноосных кристаллов.* В общем случае такое деление смысла не имеет — обе волны в кристалле ведут себя как «необыкновенные».

Чтобы выяснить физический смысл постоянных a_α, направим вектор $\mathbf{E}$ вдоль диэлектрической оси α. Тогда $\mathbf{D} = \varepsilon_\alpha\mathbf{E}$, и уравнение (75.8) перейдет в

$$v^2 = c^2\,\frac{\varepsilon_\alpha E^2}{\varepsilon_\alpha^2 E^2} = \frac{c^2}{\varepsilon_\alpha} = a_\alpha^2,$$

откуда $v = a_\alpha$. Таким образом, *величина* a_α *есть нормальная скорость распространения волны, у которой электрический вектор параллелен диэлектрической оси* α. Это утверждение становится очевидным, если заметить, что в частном случае, когда электрическое поле параллельно диэлектрической оси, уравнения распространения волн в кристалле не отличаются от уравнений в изотропных средах.

Величины a_α называются *главными скоростями распространения света в кристалле.* Наряду с главными скоростями, для характеристики оптических свойств кристаллов пользуются также *главными показателями преломления*, которые определяются выражениями

$$n_\alpha = c/a_\alpha = \sqrt{\varepsilon_\alpha}. \qquad (80.12)$$

Для волны произвольного направления показатель преломления кристалла определяется выражением

$$n = c/v. \tag{80.13}$$

Его значение, как видно из (75.8), однозначно определяется направлением вектора **D** или **E**. Каждому направлению нормали **N** соответствуют *два* значения показателя преломления в соответствии с двумя возможными поляризациями волны.

4. Для исследования уравнения Френеля применим геометрический метод. Из какой-то точки O в различных направлениях будем проводить прямые и на них откладывать отрезки, длины которых равны значениям нормальных скоростей в этих направлениях. Геометрическое место концов таких отрезков называется *поверхностью нормалей.* В кристалле каждому направлению нормали соответствуют два значения скорости. Поэтому поверхность нормалей в кристалле будет *двойной поверхностью*, т.е. состоит из двух слоев. Она представляет собой поверхность *шестого порядка* и имеет очень сложный вид. Чтобы составить представление о поверхности нормалей электромагнитных волн в кристалле, рассмотрим сечения ее координатными плоскостями XY, YZ, ZX.

Сечение плоскостью XY. Волновая нормаль лежит в плоскости XY, т.е. $N_z = 0$. Уравнение Френеля (80.8) принимает вид

$$(v^2 - a_z^2)[N_x^2(v^2 - a_y^2) + N_y^2(v^2 - a_x^2)] = 0.$$

Из него получаем два значения нормальных скоростей:

$$v_2 = a_z, \quad v_1^2 = N_x^2 a_y^2 + N_y^2 a_x^2. \tag{80.14}$$

Скорость v_2 не зависит от направления **N**. Ей соответствует круговое сечение поверхности нормалей (рис. 287). Скорость v_1 изменяется с изменением направления **N**. Ей соответствует сечение поверхности нормалей, имеющее *форму овала.* Из уравнений (80.14) следует: $v_1 \geqslant v_2$, так что круг находится целиком внутри овала. Вектор **D** должен быть перпендикулярен к **N**. Из соображений симметрии ясно, что вектор **D** одной волны параллелен оси Z, а вектор **D** другой волны параллелен плоскости XY. Первому направлению вектора **D** соответствует *круговое сечение* поверхности нормалей, второму — *овальное.*

Сечение плоскостью YZ. Волновая нормаль **N** лежит в плоскости YZ, т.е. $N_x = 0$. Уравнение Френеля принимает вид

$$(v^2 - a_x^2)[N_y^2(v^2 - a_z^2) + N_z^2(v^2 - a_y^2)] = 0.$$

Оно дает два значения нормальных скоростей:

$$v_1 = a_x, \quad v_2^2 = N_y^2 a_z^2 + N_z^2 a_y^2. \tag{80.15}$$

Скорость v_1 не зависит от направления $\mathbf{N}$. Ей соответствует *круговое сечение* поверхности нормалей и вектор $\mathbf{D}$, параллельный оси X. Скорость v_2 изменяется с изменением направления $\mathbf{N}$. Ей соответствует *овальное сечение* поверхности нормалей и вектор $\mathbf{D}$, параллельный плоскости YZ. Оно целиком помещается внутри круга, так как $v_2 \leqslant v_1$, как это следует из уравнений (80.15).

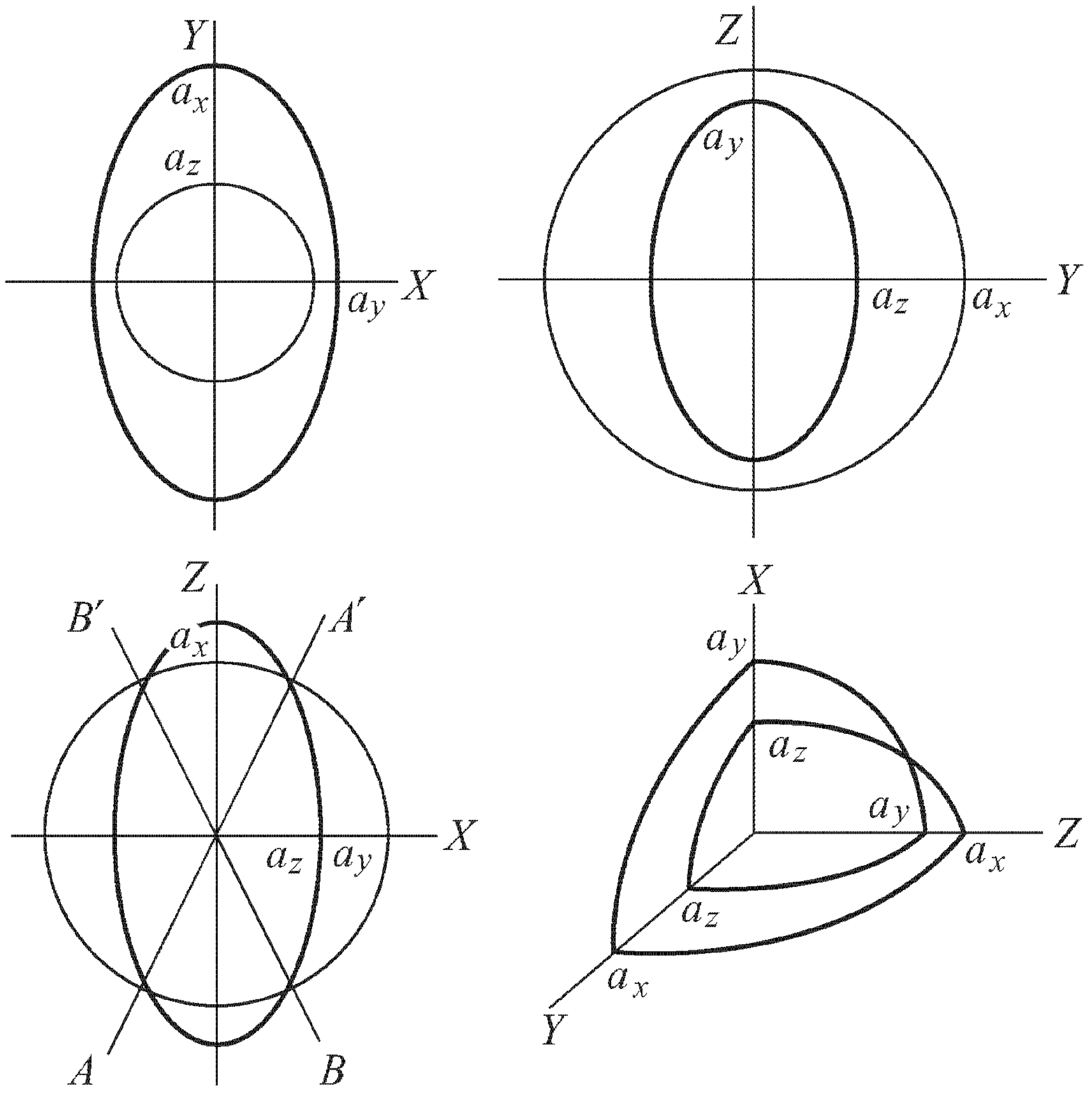


Рис. 287

Сечение плоскостью ZX. Волновая нормаль $\mathbf{N}$ лежит в плоскости ZX, т.е. $N_y = 0$. Уравнение Френеля принимает вид

$$(v^2 - a_y^2)[N_z^2(v^2 - a_x^2) + N_x^2(v^2 - a_z^2)] = 0$$

и дает два значения нормальных скоростей:

$$v_1 = a_y, \quad v_2^2 = N_z^2 a_x^2 + N_x^2 a_z^2. \tag{80.16}$$

Скорость v_1 не зависит от направления $\mathbf{N}$. Ей соответствует *круговое сечение* поверхности нормалей и вектор $\mathbf{D}$, параллельный оси Y. Скорость v_2 изменяется с изменением направления $\mathbf{N}$. Ей соответствует *овальное сечение* поверхности нормалей и вектор $\mathbf{D}$, параллельный плоскости ZX.

5. Третий из рассмотренных случаев существенно отличается от первых двух. В первых двух случаях овал и круг не пересекаются. В третьем случае они пересекаются в четырех точках (рис. 287). Это означает, что в плоскости ZX имеются *два направления* AA' и BB', симметричные относительно оси Z, вдоль которых обе волны распространяются с одной и той же нормальной скоростью. *Направления, вдоль которых совпадают нормальные скорости волн, называются оптическими осями второго рода, осями нормалей или бинормалями.*

Если в кристалле все три главные скорости a_x, a_y, a_z *различны, то в нем существуют две и только две оптические оси второго рода.* Действительно, если вектор $\mathbf{N}$ направлен вдоль оптической оси второго рода, то должно быть $v_1 = v_2$. Ввиду соотношений (80.10), это возможно только тогда, когда $v_1 = v_2 = a_y$. Но тогда уравнение (80.8) дает $N_y^2(a_y^2 - a_z^2)(a_y^2 - a_x^2) = 0$. Так как по предположению $a_x \neq a_y \neq a_z$, то отсюда следует, что $N_y = 0$. Это значит, что оптические оси лежат в плоскости ZX. Но в этой плоскости, как показано выше, имеются две и только две оптические оси второго рода. Они симметрично расположены относительно оси Z и наклонены к ней под некоторым углом β. Для нахождения β в уравнение $v_1 = v_2$ подставим значения v_1 и v_2 из формул (80.16). Получим

$$a_y^2 = N_z^2 a_x^2 + N_x^2 a_z^2, \quad \text{или} \quad a_y^2(N_x^2 + N_z^2) = N_z^2 a_x^2 + N_x^2 a_z^2,$$

откуда

$$\operatorname{tg}\beta = \frac{N_x}{N_z} = \sqrt{\frac{a_x^2 - a_y^2}{a_y^2 - a_z^2}}. \tag{80.17}$$

Теперь оправдан термин «оптически двуосный кристалл», которым мы уже пользовались.

Если две из трех главных скоростей совпадают между собой ($a_x = a_y$ или $a_y = a_z$), то оптические оси сливаются в одну ось, параллельную оси Z (когда $a_x = a_y$) или оси X (когда $a_y = a_z$). Кристалл становится *оптически одноосным.* Наконец, если все три главные скорости одинаковы, то любое направление в кристалле обладает свойством оптической оси. В таких кристаллах плоские волны, независимо от их поляризации и направления, распространяются с одной и той же скоростью — кристаллы в оптическом отношении ведут себя как *изотропные среды.* К ним относятся кристаллы кубической системы¹).

---

¹) Необходимо, однако, отметить, что при наличии *пространственной дисперсии* кристаллы кубической системы могут быть оптически анизотропными. На эту возможность указывал еще Г.А. Лорентц. Только в 1960 г. Е.Ф. Гросс и А.А. Каплянский, исследуя спектры поглощения на монокристаллических образцах Cu_2O, экспериментально обнаружили необычное для кубических кристаллов явление анизотропного поглощения света.

ЗАДАЧИ

1. Как надо ориентировать пластинку из двуосного кристалла, чтобы получить на кристалл-рефрактометре три главных показателя преломления?

Ответ. Перпендикулярно к любой из диэлектрических осей кристалла.

2. Исходя из соображений симметрии, показать, что все кристаллы три-, тетра- и гексагональной систем оптически одноосны.

§ 81. Лучи, волновые нормали и связь между ними

1. Распространение света в кристаллах, как и любых волн в анизотропных средах, характеризуется замечательной *двойственностью*, или *взаимностью*. Она обусловлена тем, что в анизотропных средах каждой волновой нормали соответствует луч, т.е. прямая, вдоль которой происходит распространение энергии волны. Поскольку энергия распространяется с групповой скоростью, для исследования свойств лучей и обоснования самого понятия луча надо вычислить групповую скорость в анизотропной среде. В этом случае такую скорость называют также *лучевой скоростью*. Для ее вычисления воспользуемся формулой (8.16), подставив в нее $\omega = kv(k)$. Дифференцируя по k_i и учитывая, что $\partial k/\partial k_i = k_i/k$, получим

$$\frac{\partial \omega}{\partial k_i} = v\frac{k_i}{k} + k\frac{\partial v}{\partial k_i}.$$

Отсюда для вектора групповой скорости находим

$$\mathbf{u} = v\mathbf{N} + k\frac{\partial v}{\partial \mathbf{k}}, \tag{81.1}$$

где $\mathbf{N} = \mathbf{k}/k$ — единичный вектор волновой нормали, а v — *нормальная скорость*, т.е. скорость распространения фазы в направлении волновой нормали.

Групповая скорость $\mathbf{u}$ в анизотропной среде отличается от нормальной скорости $\mathbf{v}$ добавочным слагаемым $k\,\partial v/\partial \mathbf{k}$. Это слагаемое в свою очередь содержит составляющую вдоль нормали $\mathbf{N}$. Чтобы определить ее, заметим, что $\mathbf{k} = k\mathbf{N}$, а потому указанная составляющая равна $\mathbf{k}\,\partial v/\partial \mathbf{k}$. Поэтому для самой групповой скорости $\mathbf{u}_N$ в направлении волновой нормали $\mathbf{N}$ можно написать

$$\mathbf{u}_N = \left(v + k\frac{\partial v}{\partial k}\right)\mathbf{N} = \left(v - \lambda\frac{\partial v}{\partial \lambda}\right)\mathbf{N}. \tag{81.2}$$

Этот результат совпадает с формулой Рэлея (8.6) для групповой скорости в изотропной среде. Этого и следовало ожидать, так как он относится не ко всему вектору групповой скорости, а только к его проекции на направление волновой нормали. А вычисление такой проекции можно провести в точности так же, как и вычисление групповой скорости в изотропной среде.

Отличие u_N от v обусловлено *дисперсией волн*, т.е. зависимостью нормальной скорости v от частоты ω. Дисперсия в равной мере свойственна и изотропным, и анизотропным средам. Специфика распространения световых волн в кристаллах обусловлена не столько дисперсией, сколько *отличием направлений волновых нормалей и лучей.* Чтобы не вводить излишних усложнений, пренебрежем совсем дисперсией, т.е. будем считать кристаллы *недиспергирующими.* Тогда $\partial v/\partial\lambda = 0$, а потому $\mathbf{u}_N = v\mathbf{N}$, или $u_N = v$. Но $u_N = (\mathbf{uN})$, так что

$$(\mathbf{uN}) = v. \tag{81.3}$$

2. Наряду с поверхностью нормалей, введенной в предыдущем параграфе, введем еще *лучевую поверхность*, называемую иногда также *волновой поверхностью.* Для этого из произвольной точки O во всевозможных направлениях будем проводить лучи и откладывать на них величины лучевой скорости в этих направлениях. Геометрическое место концов отложенных отрезков есть замкнутая поверхность, которая и называется *лучевой поверхностью.*

Если лучевую поверхность и поверхность нормалей строить из общего центра O, то между этими двумя поверхностями существует простая и важная связь. Для установления этой связи умножим формулу (81.3) на k и придадим ей вид

$$(\mathbf{uk}) = \omega. \tag{81.4}$$

Отсюда следует, что бесконечно малые изменения величин $\mathbf{u}$, $\mathbf{k}$, ω связаны соотношением $(\mathbf{u}\,\delta\mathbf{k}) + (\mathbf{k}\,\delta\mathbf{u}) = \delta\omega$. По определению групповой скорости $(\mathbf{u}\,\delta\mathbf{k}) = \delta\omega$. Следовательно, $(\mathbf{k}\,\delta\mathbf{u}) = 0$, или

$$(\mathbf{N}\,\delta\mathbf{u}) = 0. \tag{81.5}$$

Но $\mathbf{u}$ есть радиус-вектор лучевой поверхности, а потому всякий бесконечно малый вектор $\delta\mathbf{u}$ лежит в плоскости, касательной к этой поверхности в соответствующей точке касания. Поэтому из формул (81.5) и (81.3) следует, что *касательная плоскость к лучевой поверхности перпендикулярна к соответствующей волновой нормали и отсекает на ней отрезок, равный нормальной скорости волны.* Отсюда в свою очередь следует, что *лучевая поверхность есть огибающая плоских волн, распространившихся из ее центра за единицу времени в различных направлениях.* Этими теоремами и устанавливается искомая геометрическая связь между лучевой поверхностью и поверхностью нормалей.

Можно также сказать, что *касательная плоскость к лучевой поверхности есть фронт волны, соответствующий лучу, проведенному в точку касания.* В таком виде теорема допускает простую интерпретацию. Действительно, лучевая поверхность есть поверхность равных фаз, до которой световое возмущение от точечного источника доходит в течение одной секунды. Малый участок такой поверхности может рассматриваться как плоский. Если размеры участка очень велики по сравнению с длиной волны, то его распространение в течение ближайшего времени будет с достаточной точностью подчиняться законам геометрической оптики. Согласно этим законам, участок должен распространяться как безграничная плоская волна в направлении луча, причем лучевая и нормальная скорости будут связаны соотношением (81.3). Отсюда непосредственно следует, что волновой фронт есть касательная плоскость к лучевой поверхности.

Эта простая интерпретация не может, однако, заменить строгое доказательство. В ее основе лежит утверждение, что расходящийся пучок, исходящий из точечного источника, ведет себя совершенно так же, как система не зависящих друг от друга плоских волн, распространение которых чисто геометрически представляется с помощью лучевой поверхности. Впервые (1852 г.) Ламе (1795–1870) указал, что здесь необходимо решить сложную математическую задачу: точно представить волновой комплекс, исходящий в анизотропной среде из одного точечного центра (аналог шаровой волны в изотропной среде). Ламе решил эту задачу для *упругой анизотропной среды.* При этом он действительно (при исключении продольных волн) пришел к френелевой форме лучевой поверхности. В электромагнитной теории аналогичный вопрос сводится к решению задачи о поле *точечного диполя Герца, помещенного в однородную анизотропную среду.*

3. Все изложенное выше справедливо для *любых волн в анизотропных средах.* Специфичность электромагнитных волн в кристаллах состоит в том, что для них направление луча совпадает с направлением *вектора Пойнтинга.* Докажем это утверждение для рассматриваемого нами случая недиспергирующих кристаллов. В этом случае лучевая скорость вдоль волновой нормали равна $\mathbf{u}_N = v\mathbf{N} = v^2\mathbf{N}/v$, или на основании формулы (75.8)

$$\mathbf{u}_N = \frac{c^2}{D^2 v}(\mathbf{DE})\mathbf{N}. \tag{81.6}$$

Найдем теперь составляющую лучевой скорости $\mathbf{u}_\perp = k\,\partial v/\partial\mathbf{k}$, перпендикулярную к волновой нормали. Умножая формулу

(80.6) на k^2, представим ее в виде

$$\sum_\alpha \frac{k_\alpha^2}{v^2 - a_\alpha^2} = 0.$$

Дифференцируя это соотношение по k_i, получим

$$\sum_\alpha \frac{k_\alpha \partial k_\alpha / \partial k_i}{v^2 - a_\alpha^2} - \sum_\alpha \frac{k_\alpha^2}{(v^2 - a_\alpha^2)^2} v \frac{\partial v}{\partial k_i} = 0,$$

или, с учетом соотношений $\partial k_\alpha / \partial k_i = \delta_{\alpha i}$ и $k_i = kN_i$,

$$(u_\perp)_i v \sum_\alpha \frac{N_\alpha^2}{(v^2 - a_\alpha^2)^2} = \frac{N_i}{v^2 - a_i^2}.$$

Из формулы (80.5) находим

$$\frac{1}{v^2 - a_\alpha^2} = -\frac{D_\alpha}{c^2 (\mathbf{NE}) N_\alpha}, \quad \sum_\alpha \frac{N_\alpha^2}{(v^2 - a_\alpha^2)^2} = \frac{1}{c^4} \frac{D^2}{(\mathbf{NE})^2}.$$

После подстановки этих значений в предыдущее соотношение и перехода к векторной форме получим

$$\mathbf{u}_\perp = -\frac{c^2 (\mathbf{NE})}{vD^2} \mathbf{D}. \tag{81.7}$$

Следовательно,

$$\mathbf{u} = \mathbf{u}_N + \mathbf{u}_\perp = \frac{c^2}{vD^2} \{(\mathbf{DE})\mathbf{N} - (\mathbf{NE})\mathbf{D}\} = \frac{c^2}{vD^2} [\mathbf{E}[\mathbf{ND}]].$$

Согласно первой формуле (75.5),

$$[\mathbf{ND}] = -\frac{c}{v} [\mathbf{N}[\mathbf{NH}]] = \frac{c}{v} \mathbf{H},$$

а по формуле (75.8) $v^2 D^2 = c^2 (\mathbf{DE}) = c^2 H^2$, так что

$$vD = cH. \tag{81.8}$$

В результате получим

$$\mathbf{u} = \frac{c}{H^2} [\mathbf{EH}].$$

Таким образом, лучевая скорость $\mathbf{u}$, а с ней и самый луч действительно направлены вдоль вектора Пойнтинга. Так как векторы $\mathbf{E}$ и $\mathbf{H}$ взаимно перпендикулярны, то $[\mathbf{EH}] = EH\mathbf{s}$, где $\mathbf{s}$ — единичный вектор в направлении луча. С учетом этого

$$\mathbf{u} \equiv u\mathbf{s} = \frac{cE}{H} \mathbf{s}. \tag{81.9}$$

При доказательстве предполагалось, что скалярное произведение $(\mathbf{NE})$ не равно нулю. Если $(\mathbf{NE}) = 0$, то вектор $\mathbf{E}$ параллелен одной из диэлектрических осей кристалла. В этом случае теорема очевидна.

4. Обоснование понятия луча и определение его направления были даны выше через групповую скорость в анизотропной среде. Мы не воспользовались сразу теоремой Пойнтинга, чтобы показать, что это понятие и его общие свойства не специфичны для электромагнитных волн, а относятся к волнам *любой физической природы.* Кроме того, теорема Умова–Пойнтинга строго доказана для потоков энергии только через *замкнутые поверхности.* Локализация потока энергии в пространстве требует дополнительных соображений. Такая локализация не вызывает затруднений в рамках применимости геометрической оптики, т.е. в той же области, к которой относится понятие луча. Тогда теорема Умова–Пойнтинга открывает наиболее простой и общий метод для решения всех вопросов, связанных с понятием луча.

Пусть, например, AB — участок плоского волнового фронта, вырезаемый диафрагмой, поставленной на пути распространения электромагнитной волны (рис. 288). Если размеры участка велики по сравнению с длиной волны, то справедлива геометрическая оптика. Надо только определить направление световых лучей. С этой целью построим на основании AB цилиндр $ABA'B'$, образующие которого совпадают с направлениями вектора Пойнтинга. Тогда поток электромагнитной энергии через боковую поверхность цилиндра будет тождественно равен нулю. Останутся только потоки через основания AB и $A'B'$. Энергия, втекающая через AB, целиком выйдет через $A'B'$. Но так же ведут себя и световые лучи геометрической оптики. Поэтому направления лучей и вектора Пойнтинга должны совпадать и притом *не только в случае недиспергирующих сред* (как предполагалось в доказательстве, приведенном выше), *но и в случае сред, обладающих дисперсией.*

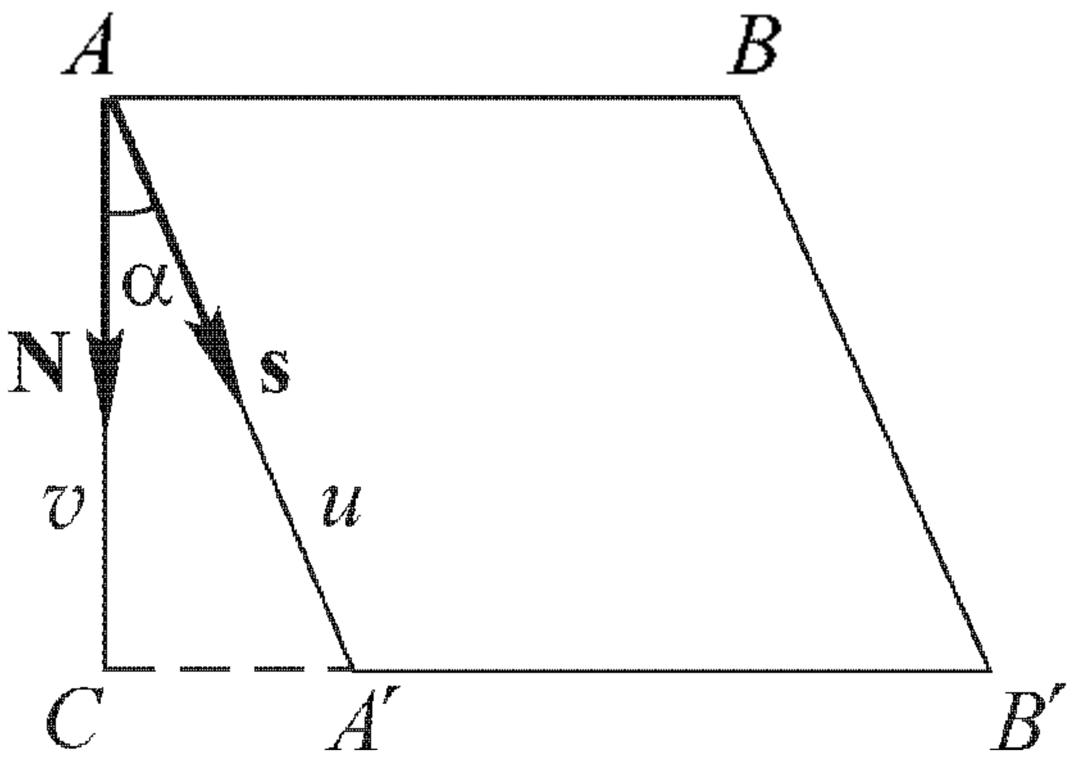


Рис. 288

5. Двойственность, характерная для кристаллооптики, о которой говорилось в начале этого параграфа, в электромагнитной теории выражается общим положением, называемым *теоремой обращения.* Эта важная теорема помогает ориентироваться в обилии сложных формул кристаллооптики, давая руководящую нить при установлении внутренних связей между ними. Выражаясь упрощенно, можно сказать, что теорема обращения сокращает вдвое число формул и теорем кристаллооптики, подлежащих запоминанию. Для вывода этой теоремы умножим первое уравнение (75.5) векторно на $\mathbf{s}$. Получим

$$[\mathbf{sD}] = -\frac{c}{v}\,[\mathbf{s}[\mathbf{NH}]] = \frac{c}{v}\,\{(\mathbf{Ns})\mathbf{H} - (\mathbf{sH})\mathbf{N}\} = \frac{c}{v}\,(\mathbf{Ns})\mathbf{H}.$$

Из (81.3) следует $u(\mathbf{Ns}) = v$. Исключая $(\mathbf{Ns})$, найдем $\mathbf{H} = \frac{u}{c}\,[\mathbf{sD}]$. Таким же путем можно получить $\mathbf{E} = -\frac{u}{c}\,[\mathbf{sH}]$. Таким образом, получаются два ряда формул:

$$\begin{aligned} \mathbf{D} &= -\frac{c}{v}\,[\mathbf{NH}], & \mathbf{E} &= -\frac{u}{c}\,[\mathbf{sH}], \\ \mathbf{H} &= \frac{c}{v}\,[\mathbf{NE}], & \mathbf{H} &= \frac{u}{c}\,[\mathbf{sD}], \\ \mathbf{N} &= \frac{1}{DH}\,[\mathbf{DH}], & \mathbf{s} &= \frac{1}{EH}\,[\mathbf{EH}], \\ D_\alpha &= \varepsilon_\alpha E_\alpha, & E_\alpha &= \frac{1}{\varepsilon_\alpha} D_\alpha. \end{aligned} \tag{81.10}$$

Все результаты, относящиеся к распространению плоских волн в однородных кристаллах, были получены нами как следствия первого ряда формул. Но формулы второго ряда можно формально получить из формул первого ряда заменой всех величин по следующей схеме:

$$\begin{array}{cccccc} \mathbf{D} & \mathbf{H} & \mathbf{N} & c & v & \varepsilon_\alpha \\ \mathbf{E} & \mathbf{H} & \mathbf{s} & 1/c & 1/u & 1/\varepsilon_\alpha \end{array} \tag{81.11}$$

Следовательно, *любое соотношение между величинами, характеризующими распространение плоских волн в однородных кристаллах, останется справедливым, если все входящие в него величины заменить на соответствующие согласно схеме* (81.11). Соответствующими считаются величины, стоящие друг под другом в рядах (81.11). Этот результат и называется *теоремой обращения.*

6. Применяя теорему обращения, легко найти лучевую скорость в кристалле, если известно направление луча $\mathbf{s}$. Для этого в уравнении (80.6) или (80.7) заменяем $\mathbf{N}$ на $\mathbf{s}$, v на $1/u$, a_α на $1/a_\alpha$ и получаем

$$\sum_\alpha \frac{s_\alpha^2}{(1/u)^2 - (1/a_\alpha)^2} = 0, \tag{81.12}$$

или

$$\sum_\alpha \frac{a_\alpha^2 s_\alpha^2}{u^2 - a_\alpha^2} = 0. \tag{81.13}$$

Это уравнение называется *законом Френеля для лучевой скорости в кристалле.* Оно вполне аналогично закону Френеля для нормальной скорости и может быть исследовано теми же способами. Но в этом нет необходимости, так как все результаты получаются непосредственно из теоремы обращения. Достаточно перечислить их.

Уравнение (81.13) *второй степени относительно* u^2. *Для каждого направления* **s** *оно имеет два положительных вещественных корня* u_1^2 *и* u_2^2, *причем*

$$a_x \geqslant u_1 \geqslant a_y \geqslant u_2 \geqslant a_z. \tag{81.14}$$

В каждом направлении в кристалле могут распространяться два линейно поляризованных луча, вообще говоря, с различными лучевыми скоростями u_1 *и* u_2. *Электрические векторы в этих двух лучах взаимно перпендикулярны.*

Лучевая поверхность, как и поверхность нормалей, состоит из *двух слоев.* Это есть поверхность *четвертого порядка.* Рассмотрим ее сечения координатными плоскостями XY, YZ и ZX. При этом можно воспользоваться прежним рис. 287, так как качественно сечения лучевой поверхности координатными плоскостями не отличаются от соответствующих сечений поверхности нормалей. Отличия, трудно передаваемые чертежом, лучше выразить словами или математическими формулами. При сечении поверхности нормалей получаются *круги* и *овалы.* Сечениями лучевой поверхности будут *круги* и *эллипсы.*

Сечение плоскостью XY. Луч **s** лежит в плоскости XY, т.е. $s_z = 0$. Лучевая скорость может иметь два значения:

$$u_2 = a_z, \qquad \frac{1}{u_1^2} = \frac{s_x^2}{a_y^2} + \frac{s_y^2}{a_x^2}. \tag{81.15}$$

Скорость u_2 не зависит от направления луча. Ей соответствует *круговое сечение* лучевой поверхности. Скорость u_1 меняется с изменением направления луча. Соответствующее сечение имеет *форму эллипса.* Действительно, уравнение рассматриваемого сечения в векторной форме имеет вид $\mathbf{r} = u_1\mathbf{s}$, откуда $s_x = x/u_1$, $s_y = y/u_1$. Подставляя эти значения в (81.15), получаем

$$\frac{x^2}{a_y^2} + \frac{y^2}{a_x^2} = 1, \tag{81.16}$$

т.е. уравнение эллипса с полуосями a_y и a_x. Ввиду соотношений (81.14) $u_1 \geqslant u_2$, так что круг находится целиком внутри эллипса. Вектор **E** должен быть перпендикулярен к **s**. Из соображений симметрии ясно, что вектор **E** одной волны параллелен оси Z, а вектор **E** другой волны параллелен плоскости XY. Первому направлению вектора **E** соответствует круговое, второму — эллиптическое сечение лучевой поверхности.

Сечение плоскостью YZ. В этой плоскости лучевые скорости могут иметь два значения:

$$u_1 = a_x, \qquad \frac{1}{u_2^2} = \frac{s_y^2}{a_z^2} + \frac{s_z^2}{a_y^2}. \tag{81.17}$$

Скорость u_1 не зависит от направления луча. Ей соответствует круговое сечение лучевой поверхности и вектор **E**, параллельный оси X. Скорость u_2 меняется с изменением направления луча. Соответствующее сечение есть эллипс

$$\frac{y^2}{a_z^2} + \frac{z^2}{a_y^2} = 1, \tag{81.18}$$

а вектор $\mathbf{E}$ лежит в плоскости ZX. Согласно (81.14) $u_1 \geqslant u_2$, так что эллипс целиком помещается внутри круга.

Сечение плоскостью ZX. Скорость луча определяется двумя выражениями:

$$u_1 = a_y, \qquad \frac{1}{u_2^2} = \frac{s_z^2}{a_x^2} + \frac{s_x^2}{a_z^2}. \tag{81.19}$$

Скорости u_1 соответствует круговое сечение лучевой поверхности и вектор $\mathbf{E}$, параллельный оси Y. Для u_2 получается эллипс

$$\frac{z^2}{a_x^2} + \frac{x^2}{a_z^2} = 1, \tag{81.20}$$

а вектор $\mathbf{E}$ лежит в плоскости ZX. Эллипс и круг пересекаются друг с другом в четырех точках (рис. 287). В соответствии с этим в плоскости ZX имеются два направления AA' и BB', симметричные относительно оси Z, вдоль которых оба луча распространяются с одинаковыми лучевыми скоростями. Такие направления называются *оптическими осями первого рода*, *лучевыми осями* или *бирадиалями*.

Если в кристалле все три главные скорости a_x, a_y, a_z различны, то в нем существуют две и только две оптические оси первого рода. Они лежат в плоскости ZX и симметрично расположены относительно оси Z. Угол γ, образуемый одной из оптических осей первого рода с осью Z, определяется формулой

$$\operatorname{tg}\gamma = \frac{s_x}{s_z} = \sqrt{\frac{(1/a_x^2) - (1/a_y^2)}{(1/a_y^2) - (1/a_z^2)}} = \frac{a_z}{a_x}\sqrt{\frac{a_x^2 - a_y^2}{a_y^2 - a_z^2}}. \tag{81.21}$$

Сравнение этой формулы с формулой (80.17) приводит к соотношению

$$\operatorname{tg}\gamma = \frac{a_z}{a_x}\operatorname{tg}\beta, \tag{81.22}$$

из которого следует: $\gamma < \beta$, т.е. *оптические оси первого рода расположены ближе к оси Z, чем оптические оси второго рода*. Обычно a_x и a_z не очень сильно отличаются друг от друга. Поэтому угол между оптическими осями первого и второго рода, как правило, мал и при рассмотрении многих явлений может не приниматься во внимание. Для слюды он составляет около $40'$.

Если две из трех главных скоростей равны между собой, то оптические оси второго рода сливаются в одну ось, направленную либо параллельно оси Z (когда $a_x = a_y$), либо параллельно оси X (когда $a_y = a_z$). В этом случае оптическая ось первого рода совпадает с оптической осью второго рода. Наконец, когда все три главные скорости равны между собой, любое направление в кристалле обладает свойствами оптической оси.

По числу оптических осей первого рода кристаллы разделяются на: 1) *двуосные*, 2) *одноосные* и 3) *оптически изотропные*. Эта классификация совпадает с классификацией, основанной на числе оптических осей второго рода.

7. Отметим еще одно следствие теоремы обращения, которое понадобится нам в следующем параграфе. Умножая скалярно уравнение (75.7) на $\mathbf{s}$ и принимая во внимание, что $(\mathbf{Es}) = 0$, получим

$$(\mathbf{Ds}) = -\frac{c^2}{v^2}(\mathbf{NE})(\mathbf{Ns}) = -\frac{c^2}{uv}(\mathbf{NE}). \tag{81.23}$$

Преобразуем (80.3) с помощью теоремы обращения:

$$\left(\frac{1}{u^2} - \frac{1}{a_\alpha^2}\right) E_\alpha = -\frac{1}{c^2}(\mathbf{sD})s_\alpha,$$

или на основании (81.23)

$$\frac{a_\alpha^2 - u^2}{a_\alpha^2} E_\alpha = \frac{u}{v}(\mathbf{NE})s_\alpha.$$

Умножим и разделим левую часть этого соотношения на ε_α и учтем, что $D_\alpha = \varepsilon_\alpha E_\alpha$, $a_\alpha^2 \varepsilon_\alpha = c^2$. Тогда

$$(a_\alpha^2 - u^2)D_\alpha = \frac{c^2 u}{v}(\mathbf{NE})s_\alpha. \tag{81.24}$$

Сравнение этого соотношения с (80.3) дает

$$u(a_\alpha^2 - v^2)s_\alpha = v(a_\alpha^2 - u^2)N_\alpha, \tag{81.25}$$

откуда

$$\frac{u s_\alpha}{a_\alpha^2 - u^2} = \frac{v N_\alpha}{a_\alpha^2 - v^2}. \tag{81.26}$$

Умножая это соотношение на N_α, суммируя по α и принимая во внимание (80.6), получим

$$\sum_\alpha \frac{N_\alpha s_\alpha}{a_\alpha^2 - u^2} = 0, \tag{81.27}$$

или

$$N_x s_x(a_y^2 - u^2)(a_z^2 - u^2) + N_y s_y(a_z^2 - u^2)(a_x^2 - u^2) + \\ + N_z s_z(a_x^2 - u^2)(a_y^2 - u^2) = 0$$

— соотношение более общее, чем (81.27), так как справедливость его, как легко показать, не связана с предположением, что $a_\alpha^2 - u^2 \neq 0$.

8. Теорема о связи между лучевой поверхностью и поверхностью нормалей (см. п. 2) позволяет геометрически построить одну из этих поверхностей, если известна другая. Пусть, например, ACB — участок лучевой поверхности с центром O (рис. 289). В каждой точке этой лучевой поверхности проведем касательную плоскость и опустим на нее перпендикуляр из центра O. Геометрическое место оснований таких перпендикуляров будет поверхностью нормалей. Наоборот, чтобы по заданной поверхности нормалей построить лучевую поверхность, надо из центра O провести во всевозможных направлениях радиусы-векторы и в точках пересечения их с поверхностью нормалей построить плоскости, перпендикулярные к ним. Огибающая таких плоскостей и будет лучевой поверхностью.

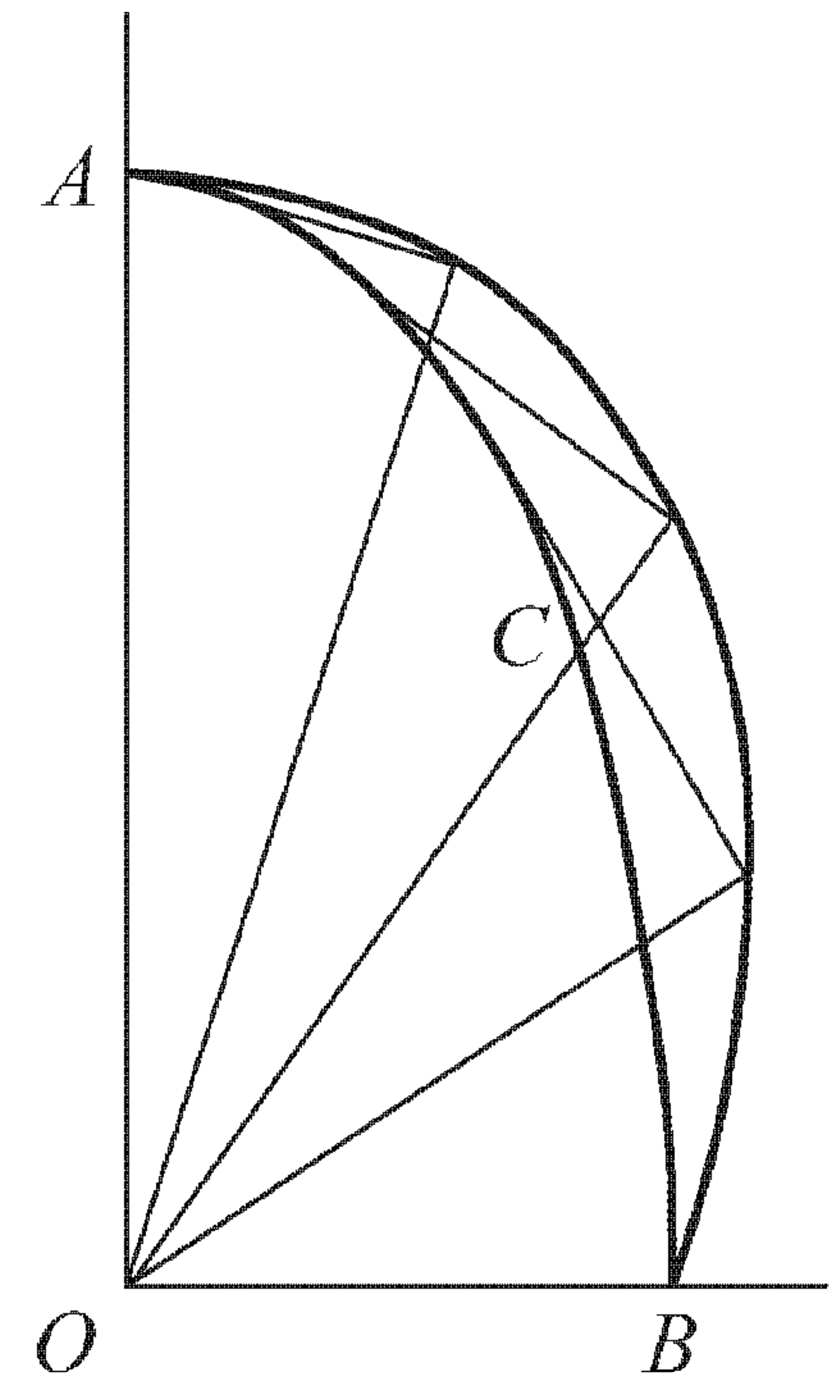


Рис. 289

Допустим, что в некоторый момент времени t в кристалле известно положение плоского волнового фронта. Для того чтобы построить волновой фронт в более поздний момент времени t', можно на основании доказанной теоремы поступить следующим образом. Из каждой точки исходного волнового фронта опишем элементарную волну, радиусы-векторы которой получаются умножением на $(t' - t)$ соответствующих радиусов-векторов лучевой поверхности. Плоскость, касательная ко всем элементарным волнам, и даст положение волнового фронта в момент времени t'. Из двух возможных касательных плоскостей следует выбрать ту, которой соответствует волна требуемой поляризации. Направление луча найдется соединением центра элементарной волны с соответствующей точкой касания.

Это построение является обобщением *построения Гюйгенса* для изотропных сред. Оно было впервые введено Гюйгенсом для объяснения двойного преломления. Гюйгенс постулировал, что элементарная волна в кристаллах состоит из двух волн: *сферической* и *эллипсоидальной*. Сферические волны порождают *обыкновенную*, а эллипсоидальные — *необыкновенную волны*. Это предположение Гюйгенса оправдалось, но оно верно только для *оптически одноосных кристаллов*.

§ 82. Коническая рефракция

1. Когда волновая нормаль $\mathbf{N}$ параллельна одной из оптических осей второго рода, нормальные скорости обеих волн v_1 и v_2 совпадают между собой, а направления векторов $\mathbf{D}$ становятся неопределенными. Значит, в направлении оптической оси второго рода может распространяться плоская волна любой поляризации, причем скорость распространения не зависит от характера поляризации. В этом отношении рассматриваемый случай аналогичен распространению волн в изотропной среде. Однако если кристалл двуосный, между ними имеется существенное различие.

В изотропной среде направления векторов **D** и **E**, а также **N** и **S** всегда совпадают. Для волны, распространяющейся вдоль оптической оси двуосного кристалла, положение меняется. В этом случае вектор **D** может принимать любое направление, перпендикулярное к **N**. Так как **N** лежит в плоскости ZX, то одним из возможных направлений вектора **D** будет ось Y. Тогда и только тогда вектор **D** будет совпадать по направлению с вектором **E**, а нормаль **N** — с лучом **s**. Во всех остальных случаях направления луча и волновой нормали отличаются друг от друга. Придавая **D** всевозможные направления, перпендикулярные к оптической оси, получим бесчисленное множество направлений луча **s**. Направление вектора **s** становится *неопределенным*. Докажем, что в рассматриваемом случае все лучи лежат на поверхности конуса.

Для доказательства воспользуемся соотношением (81.27), которое справедливо при любом направлении волновой нормали. Так как волновая нормаль **N** направлена вдоль оптической оси, то $N_y = 0$, и соотношение (81.27) принимает вид

$$\frac{N_x s_x}{a_x^2 - u^2} + \frac{N_z s_z}{a_z^2 - u^2} = 0.$$

Отсюда

$$a_z^2 N_x s_x + a_x^2 N_z s_z = u^2 (\mathbf{N}\mathbf{s}).$$

Ввиду соотношения (81.3), $u(\mathbf{N}\mathbf{s}) = v = a_y$. Следовательно,

$$(a_z^2 N_x s_x + a_y^2 N_z s_z)(N_x s_x + N_z s_z) = a_y^2.$$

Рассмотрим произвольную точку на луче **s** с радиусом-вектором $\mathbf{r}(x, y, z)$. Очевидно, $x_\alpha = r s_\alpha$, и предыдущее соотношение переходит в

$$(a_z^2 N_x x + a_y^2 N_z z)(N_x x + N_z z) = a_y^2 (x^2 + y^2 + z^2). \qquad (82.1)$$

Это однородное уравнение второго порядка представляет *конус*. Образующими конуса являются лучи, соответствующие волновой нормали **N**, параллельной одной из двух оптических осей второго рода. Конус (82.1) называется *конусом внутренней конической рефракции*. Волновая нормаль есть одна из образующих конуса (82.1). Это следует из того, что направления **s** и **N** совпадают, когда вектор **D** параллелен диэлектрической оси Y.

2. Конус внутренней конической рефракции пересекается фронтом волны

$$\mathbf{N}\mathbf{r} \equiv N_x x + N_z z = a_y \qquad (82.2)$$

по кругу. В самом деле, линия пересечения определяется системой уравнений (82.1) и (82.2), равносильной системе

$$\begin{aligned} a_z^2 N_x x + a_y^2 N_z z &= a_y (x^2 + y^2 + z^2), \\ \mathbf{N}\mathbf{r} &= a_y. \end{aligned} \qquad (82.3)$$

Первое уравнение есть уравнение сферы, второе — уравнение плоскости. Их пересечение есть круг, и наше утверждение доказано.

Определим угол раствора конуса внутренней конической рефракции, точнее — угол χ, получающийся от пересечения этого конуса плоскостью ZX, проходящей через оптическую ось кристалла (рис. 290). Когда вектор $\mathbf{D}$ направлен вдоль диэлектрической оси Y, векторы $\mathbf{s}$ и $\mathbf{N}$, а также оптическая ось второго рода совпадают по направлению. Если же вектор $\mathbf{D}$ лежит в плоскости ZX, то в той же плоскости будет лежать и луч $\mathbf{s}$, так как четыре вектора $\mathbf{E}$, $\mathbf{D}$, $\mathbf{s}$, $\mathbf{N}$ всегда должны лежать в одной плоскости (см. § 75). Искомый угол χ будет равен углу между векторами $\mathbf{D}$ и $\mathbf{E}$, поскольку стороны этих углов взаимно перпендикулярны. Его легко определить из формулы (75.8), так как в рассматриваемом случае нормальная скорость v равна a_y. Формула (75.8) дает

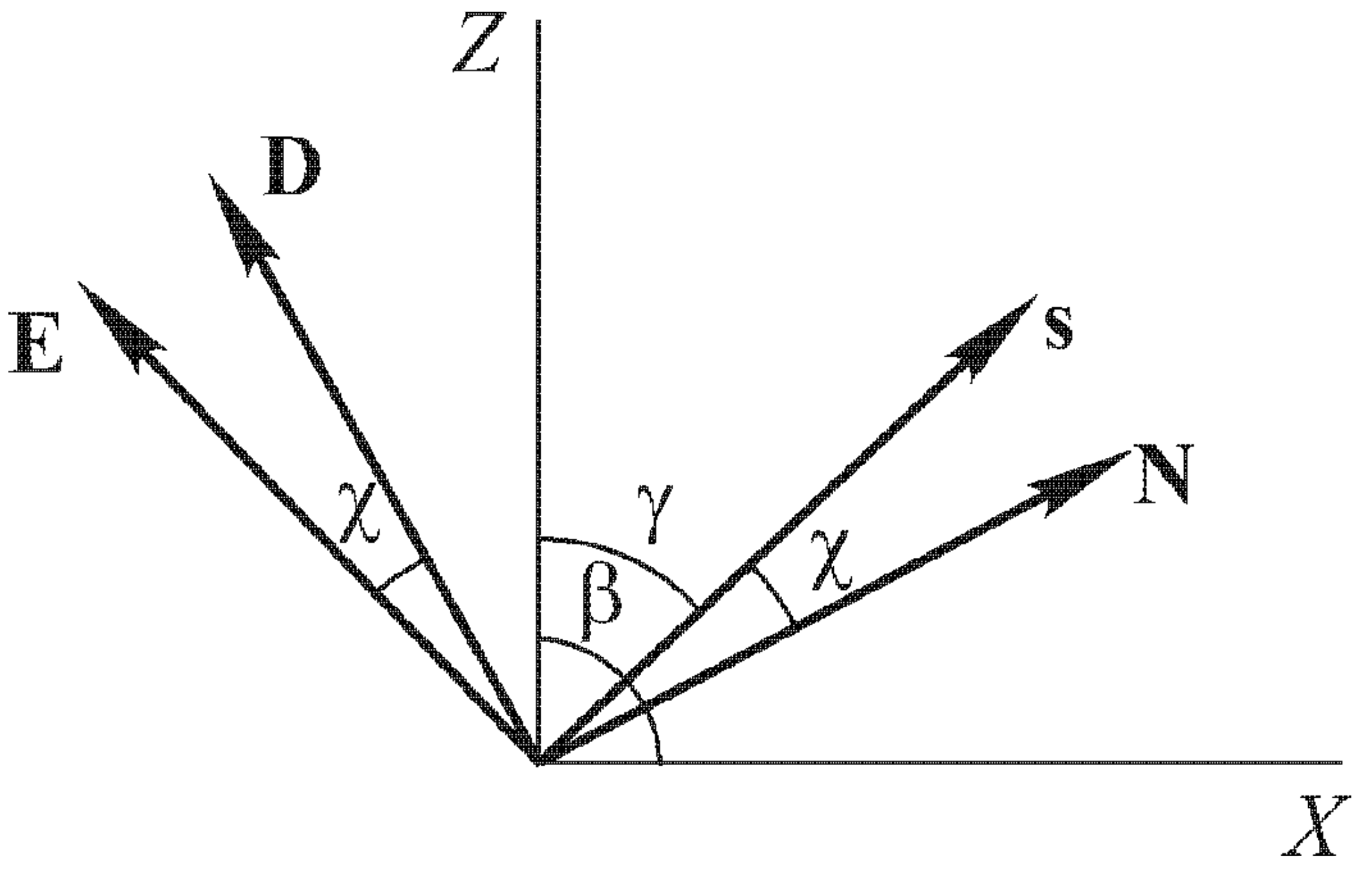


Рис. 290

$$\cos\chi = \frac{D}{E}\left(\frac{a_y}{c}\right)^2.$$

Как видно из рис. 290, $D_x = -D\cos\beta$, $D_z = D\sin\beta$, где β — угол между оптической осью второго рода и осью Z. Поэтому

$$E_x = \frac{D_x}{\varepsilon_x} = -\frac{D\cos\beta}{c^2}a_x^2, \qquad E_z = \frac{D_z}{\varepsilon_z} = \frac{D\sin\beta}{c^2}a_z^2.$$

Подставляя эти значения в выражение для $\cos\chi$, получим

$$\cos\chi = \frac{a_y^2}{\sqrt{a_x^4\cos^2\beta + a_z^4\sin^2\beta}},$$

откуда

$$\operatorname{tg}\chi = \frac{1}{a_y^2}\sqrt{a_x^4\cos^2\beta + a_z^4\sin^2\beta - a_y^4}.$$

Использовав формулу (80.17), после несложных преобразований найдем

$$\operatorname{tg}\chi = \frac{1}{a_y^2}\sqrt{(a_x^2 - a_y^2)(a_y^2 - a_z^2)}. \qquad (82.4)$$

Конус внутренней конической рефракции пересекает лучевую поверхность по кругу, вдоль которого ее касается фронт волны. Это непосредственно следует из теоремы, доказанной в § 81 (п. 2).

3. Теорема обращения распространяет полученные результаты на лучи. Если луч в двуосном кристалле направлен вдоль одной из оптических осей первого рода, то ему соответствует бесконечное множество волновых нормалей, образующих конус. Этот конус называется *конусом внешней конической рефракции.* Луч есть одна из образующих этого конуса. Сечение конуса внешней конической рефракции плоскостью, перпендикулярной лучу, есть круг. Угол раствора конуса определяется уравнением

$$\operatorname{tg}\psi = a_y^2\sqrt{\left(\frac{1}{a_x^2} - \frac{1}{a_y^2}\right)\left(\frac{1}{a_y^2} - \frac{1}{a_z^2}\right)} = \frac{1}{a_x a_z}\sqrt{(a_x^2 - a_y^2)(a_y^2 - a_z^2)}. \qquad (82.5)$$

Проведем касательную плоскость к лучевой поверхности в точке S пересечения ее с лучевой осью. Такая плоскость будет перпендикулярна к волновой нормали. А так как волновых нормалей, соответствующих лучу, направленному вдоль лучевой оси, бесконечно много, то в точке S можно провести бесконечное множество касательных плоскостей к лучевой поверхности. Это означает, что в окрестности такой точки лучевая поверхность имеет воронкообразную форму.

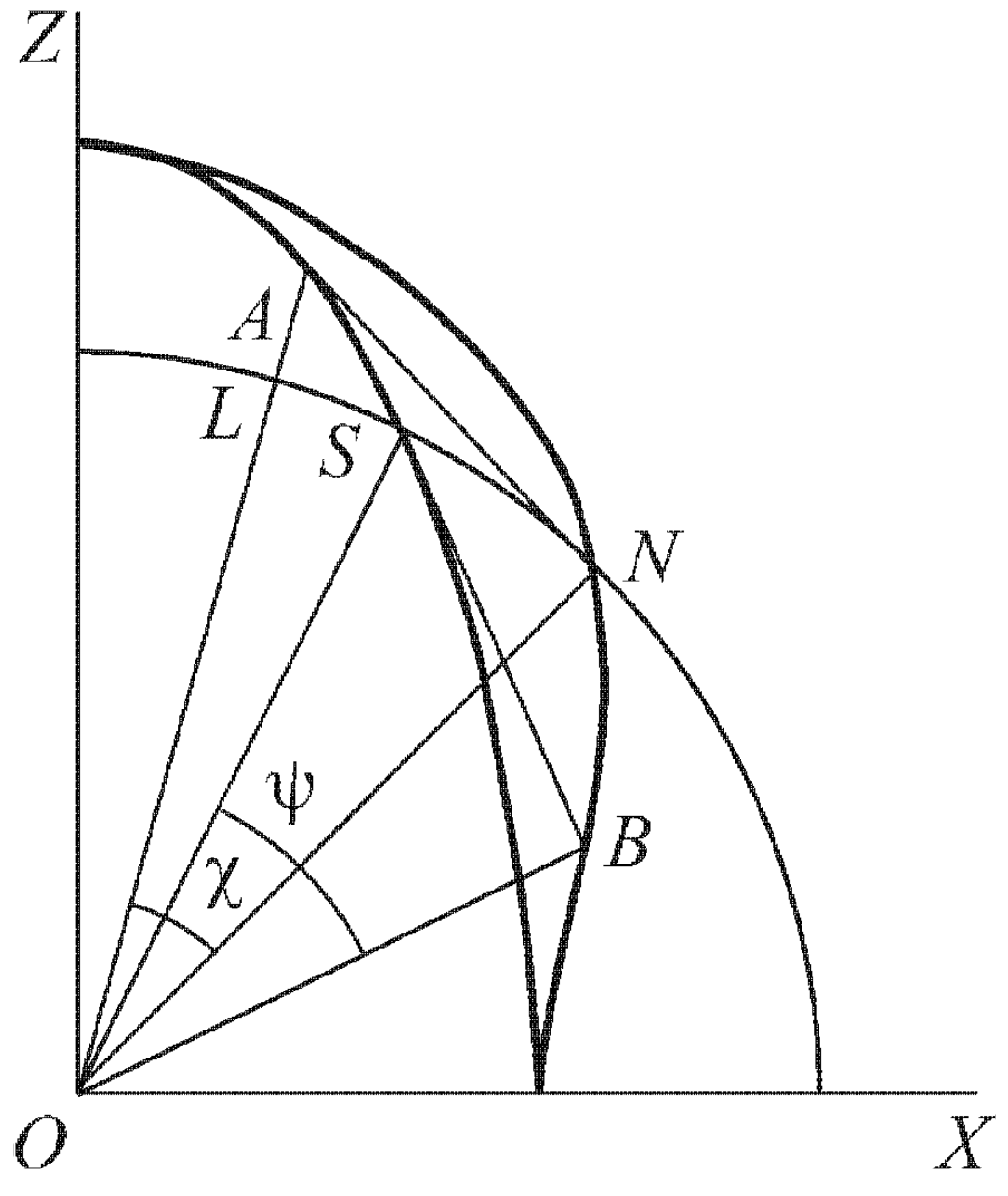


Рис. 291

На рис. 291 представлено сечение поверхности нормалей и лучевой поверхности плоскостью ZX. Точка N есть двойная точка поверхности нормалей, ON — оптическая ось второго рода. Перпендикуляр NA к этой оси дает сечение фронта волны плоскостью рисунка. Прямая NA касается лучевой поверхности в точке A, угол $\chi = \angle NOA$ есть угол раствора конуса внутренней конической рефракции, S — двойная точка лучевой поверхности, OS — лучевая ось. Касательная к лучевой поверхности в точке S пересекает поверхность нормалей в точке B; прямая OB будет одной из волновых нормалей, принадлежащих лучу OS. Сам луч OS является нормалью плоской волны, которая касается кругового сечения лучевой поверхности $u = a_y$ в точке S. Угол $\psi = \angle SOB$ есть угол раствора конуса внешней конической рефракции.

Каждому лучу, принадлежащему конусу внутренней конической рефракции, например лучу OA (рис. 291), соответствует вполне определенная линейная поляризация. В самом деле, в направлении OA могут распространяться два луча, электрические векторы которых взаимно перпендикулярны. Однако только один из них соответствует волне, распространяющейся вдоль волновой нормали ON. Другому лучу соответствует лучевая скорость OL и, следовательно, иное направление волновой нормали. Аналогично, каждой волновой нормали, принадлежащей конусу внешней конической рефракции, также соответствует вполне определенная линейная поляризация.

4. Волновой фронт AN, распространяющийся в направлении оптической оси второго рода ON, как было показано, касается лучевой поверхности по кругу, вдоль которого эта поверхность пересекается конусом внутренней конической рефракции. Такой волновой фронт не может пересекать лучевую поверхность. В самом деле, пересечем лучевую поверхность плоскостью ANO, проходящей через оптическую ось ON (рис. 292). В сечении получится кривая CBA. Если бы фронт AN пересекал лучевую поверхность, то, ввиду конечности лучевой скорости, на кривой CBA нашлись бы такие точки B, C, что проекции радиусов-векторов OB, OC на направление ON были бы минимальны или максимальны. Если плоскость ANO проводить во всевозможных направлениях, проходящих через оптическую ось ON, то точки

Рис. 292

B, C опишут замкнутые кривые. На этих кривых в свою очередь найдутся точки, проекции радиусов-векторов которых на направление бинормали ON будут максимальны или минимальны. Пусть B — одна из таких точек. Тогда касательная плоскость BN' к лучевой поверхности в точке B будет параллельна волновому фронту AN, т.е. перпендикулярна к бинормали ON. Значит, плоскость BN' сама является волновым фронтом, распространяющимся в направлении бинормали ON.

Таким образом, если бы волновой фронт NA мог пересекать лучевую поверхность, то вдоль бинормали могли бы распространяться две волны с различными нормальными скоростями $v = ON$ и $v' = ON'$, что противоречит определению бинормали. Поскольку волновой фронт касается лучевой поверхности по кругу (в точках A), из доказанного следует, что лучевая поверхность целиком лежит с той стороны волнового фронта AN, с которой находится ее центр O (рис. 291). Значит, в точке S лучевая поверхность имеет не просто воронкообразную форму, как было отмечено выше, но *воронкообразное углубление.*

5. После выяснения этих геометрических соотношений обратимся к рассмотрению внутренней конической рефракции, теоретически предсказанной Гамильтоном (1805–1865) в 1832 г. Примерный ход рассуждений Гамильтона был следующий. Пусть плоскопараллельная пластинка из двуосного кристалла прикрыта с одной стороны непрозрачным экраном с малым отверстием O (рис. 293). Осветим пластинку параллельным пучком неполяризованных лучей таким образом, чтобы после преломления на передней поверхности пластинки волновая нормаль оказалась направленной вдоль одной из оптических осей второго рода OA. Волновой нормали OA соответствует конус лучей. Энергия распространяется вдоль лучей, поэтому при достаточно малых размерах отверстия O световой пучок внутри пластинки развернется в конус OAB. После преломления на задней поверхности пластинки волновая нормаль примет свое исходное направление. А так как в изотропных средах направления лучей и волновых нормалей совпадают, то все лучи выйдут из пластинки параллельным пучком и расположатся по поверхности цилиндра. Если на их пути поместить экран, то на нем должно получиться светлое кольцо. По предложению Гамильтона отысканием этого явления занялся Ллойд, который и обнаружил его в 1833 г. на кристалле арагонита: на экране наблюдалось светлое эллиптическое кольцо. (Угол χ для арагонита равен $1°52'$.)

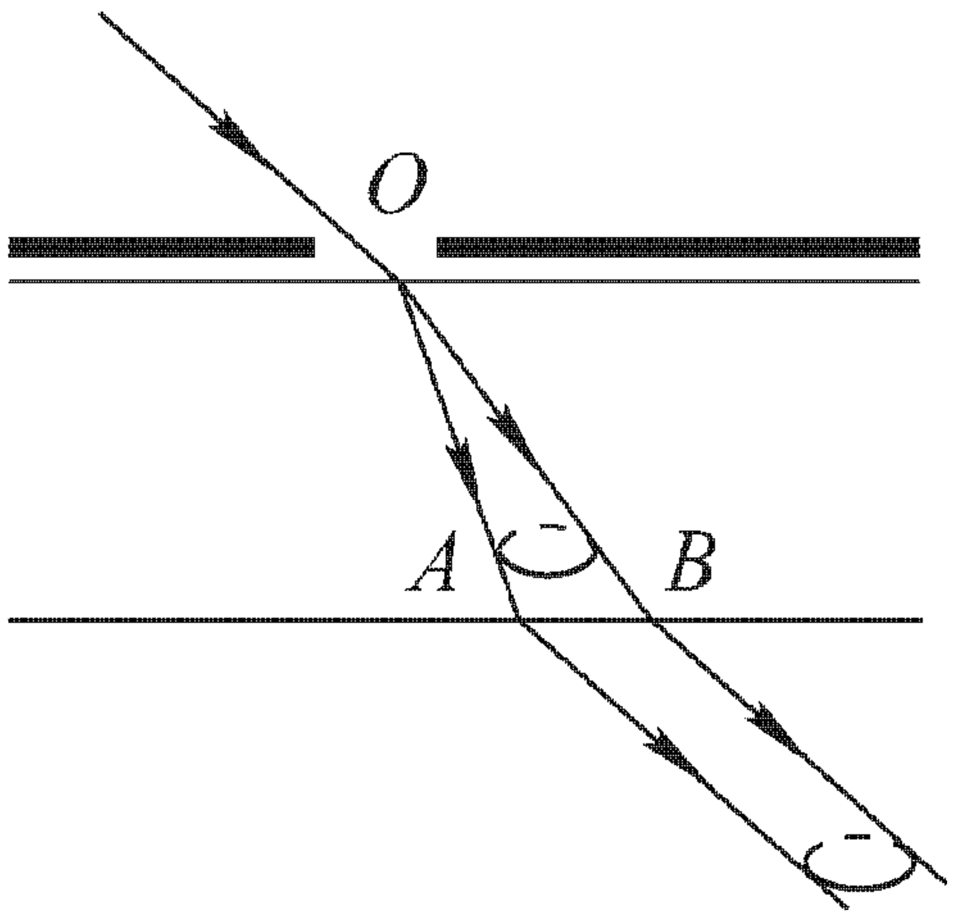


Рис. 293

Хотя Гамильтон и предсказал коническую рефракцию, его объяснение неправильно. При более детальном изучении оказалось, что явление выглядит иначе, чем предсказывал Гамильтон. Применяя более узкие отверстия в экране, Поггендорф (1796–1877) нашел, что кольцо в действительности *двойное.* Объяснение было дано Фохтом (1850–1919). Гамильтон рассматривал строго плоскую волну, распространяющуюся в кристалле точно в направлении оптической оси. Физически это реализовать невозможно. Если бы даже можно было осветить отверстие O строго плоской волной, то после прохождения через него волна перестала бы быть плоской из-за дифракции. Такая волна распадается на бесконечное множество плоских волн, направления распространения которых близки к направлению оптической оси. Нельзя ограничиться рассмотрением поведения только одной волны, распространяющейся строго в направлении оптической оси. Это ясно уже из того, что на ее долю приходится исчезающе малая энергия, и физически ничего не изменится, если эту волну даже совсем удалить из волнового комплекса. Необходимо рассмотреть бесконечное множество плоских волн, волновые нормали которых группируются вблизи оптической оси. Это и было сделано Фохтом.

Если строго плоская волна распространяется в направлении оптической оси второго рода, то, как было показано, волновой фронт касается лучевой поверхности по кругу. Примем плоскость такого круга за плоскость рис. 294. Пусть N — точка пересечения плоскости рисунка с оптической осью. Допустим, что имеется бесконечная совокупность плоских волн, волновые нормали которых лежат в пределах небольшого конуса, ось которого совпадает с оптической осью. Этот конус пересекается плоскостью рисунка по кругу k. Каждой волновой нормали соответствует точка внутри или на границе круга k. Найдем, где расположатся соответствующие ей лучи. Волновой нормали N соответствует конус лучей, пересекающих плоскость рисунка по окружности K. Лучевая поверхность касается плоскости рисунка вдоль этой окружности, а поэтому кривизна лучевой поверхности в направлении окружности K равна нулю. Кроме того, лучевая поверхность должна лежать по одну сторону от плоскости рисунка. Ради определенности примем, что все волны распространяются к читателю. Тогда лучевая поверхность будет лежать за плоскостью рисунка, и следовательно, ее окрестность вблизи окружности K будет обращена к читателю своей выпуклостью. В такой окрестности лучевая поверхность имеет баранкообразную форму.

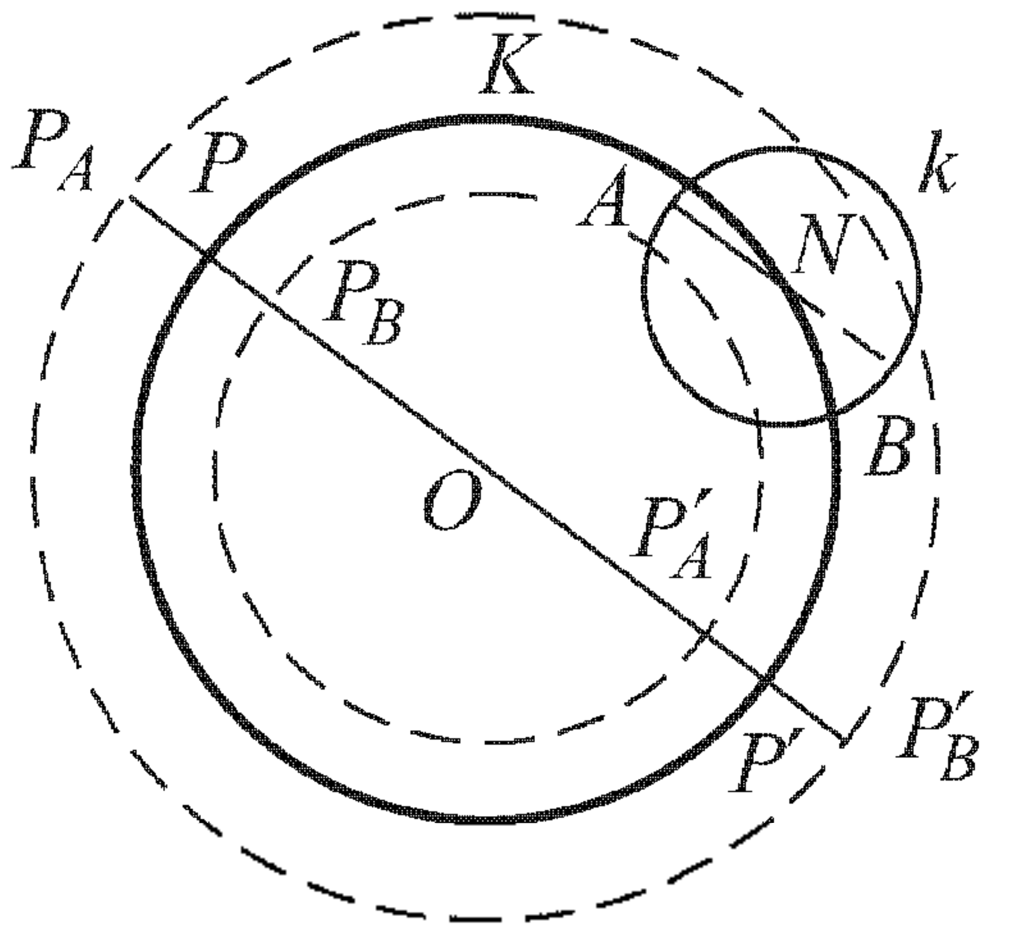


Рис. 294

Пересечем лучевую поверхность двумя параллельными плоскостями P_AP_B и AB, перпендикулярными к плоскости рисунка и проходящими через центры окружностей K и k. Бесконечно малые отрезки P_AP_B и $P'_AP'_B$ перпендикулярны к окружности K, поэтому в направлениях этих отрезков кривизна лучевой поверхности будет максимальна, а в перпендикулярных направлениях равна нулю. Следовательно, перпендикуляры к этим бесконечно малым отрезкам должны лежать в плоскости $P_AP'_B$, т.е. они будут параллельны волновым нормалям, лежащим в плоскости AB. Значит, касательная плоскость к лучевой поверхности в какой-либо точке отрезка P_AP или отрезка P'_AP' будет перпендикулярна к соответствующей волновой нормали, проходящей через отрезок AN. Касательная же плоскость к лучевой поверхности в какой-либо точке отрезков PP_B в $P'P'_B$ будет перпендикулярна к волновой нормали, проходящей через отрезок NB. Это означает, что каждой волновой нормали, проходящей через отрезок AN, соответствуют два луча, из которых один проходит через отрезок P_AP, а другой — через отрезок P'_AP'. Каждой же волновой нормали, проходящей через отрезок NB, соответствуют два луча, проходящие через отрезки PP_B и $P'P'_B$. Таким образом, каждой волновой нормали, наклоненной под малым углом к оптической оси второго рода, соответствуют два луча, один из которых проходит внутри конуса внутренней конической рефракции, а другой вне этого конуса.

Теперь ясно происхождение двойного светлого кольца Поггендорфа. Волновым нормалям, пересекающим плоскость рисунка внутри малого круга радиуса dr с центром в N, соответствует малая доля энергии, которая должна распределиться по сравнительно большой площади dS кольца по обе стороны от окружности K. Если же взять волновые нормали, пересекающие плоскость рисунка внутри кольца со средним радиусом r и той же толщиной dr, то таким волновым нормалям будет соответствовать значительно бóльшая энергия, поскольку она пропорциональна площади кольца $2\pi r\,dr$. Эта энергия должна распределиться по площади двух колец, одно из которых лежит внутри, а другое вне окружности K. Площади обоих колец с точностью до бесконечно малых высшего порядка по-прежнему равны dS. Поэтому освещенность обоих колец будет много больше освещенности центрального кольца в окрестности окружности K. Освещенность должна равняться нулю вдоль окружности K и непрерывно возрастать по мере удаления от этой окружности как в наружную, так и во внутреннюю стороны. Таким образом, там, где по Гамильтону должна была бы получаться максимальная освещенность, в действительности наблюдается темнота.

6. Гамильтоном была предсказана еще *внешняя коническая рефракция*, экспериментально обнаруженная также Ллойдом в 1833 г. Она связана с тем, что световому лучу, идущему вдоль лучевой оси двуосного кристалла, соответствует бесконечная совокупность волновых нормалей, лежащих на конической поверхности. Нет необходимости входить в теорию этого явления. Достаточно сослаться на теорему обращения и описать явление, как оно наблюдалось в установке Ллойда. Обе поверхности плоскопараллельной арагонитовой пластинки были покрыты экранами с малыми отверстиями O и O' (рис. 295), центры которых лежали на оптической оси первого рода. Линза L концентрировала на O сходящийся пучок лучей. Диафрагмы O и O' выделяли только те лучи, которые шли вдоль оптической оси OO'. По выходе из пластинки лучи развертывались в конус, и на экране Э наблюдалось светлое кольцо. Разумеется, при достаточно малых размерах отверстий O и O' кольцо будет *двойное*.

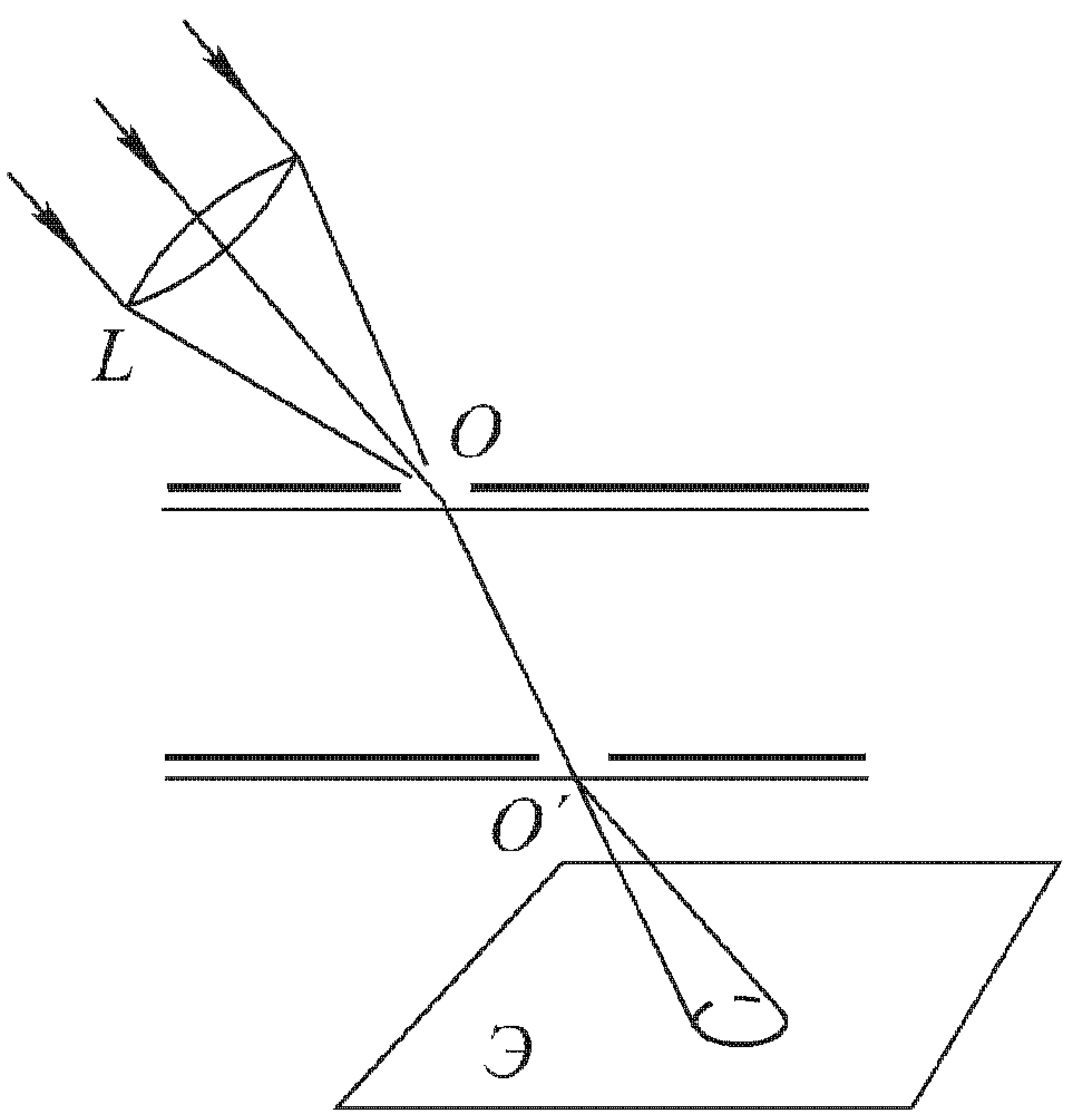


Рис. 295

§ 83. Замечания об отражении и преломлении света на границе кристаллов

1. Как и в оптике изотропных сред, задача об отражении и преломлении в кристаллооптике может быть полностью решена на основе *граничных условий*, которым должны удовлетворять векторы электромагнитного поля: по обе стороны границы раздела тангенциальные компоненты векторов **E** и **H** должны быть равны. Задача распадается на две части: *геометрическую*, в которой определяются направления распространения, поляризация и скорости отраженных и преломленных волн, а в случае их неоднородности также и затухание волн в пространстве, и *физическую*, в которой определяются амплитуды волн (вообще говоря, комплексные). Остановимся кратко только на геометрической стороне вопроса. Формулы для амплитуд, из-за их громоздкости, выводить не будем, хотя сам вывод и не встречает принципиальных затруднений.

Пусть на плоскую границу раздела падает плоская монохроматическая волна с волновым вектором $\mathbf{k}_1$. В случае изотропных сред получается только одна отраженная и только одна преломленная волна. Для анизотропных сред это, вообще говоря, не так. Однако, каково бы ни было число отраженных и преломленных волн, из линейности и однородности граничных условий непосредственно следует, что тангенциальные компоненты волновых векторов падающей, отраженных и преломленных волн должны быть одинаковы (см. § 69). Следовательно, *нормали падающей, отраженных и преломленных волн, а также нормаль к границе раздела все лежат в одной плоскости.* Кроме того, *преломление волновых нормалей подчиняется закону преломления Снеллиуса: отношение синуса угла падения к синусу угла преломления равно отношению соответствующих нормальных скоростей волн.* Практически от этого закона мало пользы. Его недостаточно для нахождения направления волновой нормали преломленной волны, так как нормальная скорость преломленной волны *неизвестна.* Она зависит от направления падающей волны и *сама подлежит определению.* Проще всего это можно сделать графическим методом.

2. Рассмотрим одну из преломленных волн. Пусть $\mathbf{k} = \overrightarrow{OA}$ или $\mathbf{k} = \overrightarrow{OB}$ означает ее волновой вектор. Ввиду граничных условий, тангенциальную составляющую этого вектора $\mathbf{k}_t$ можно считать известной: $\mathbf{k}_t = \mathbf{k}_{1t}$. Она лежит в плоскости падения. Остается найти нормальную составляющую $\mathbf{k}_n$. Прежде всего, она должна быть направлена вниз, в сторону второй среды, так как преломленная волна должна *уходить от границы раздела.* Для решения задачи можно применить следующее геометрическое построение. Будем проводить из точки O (рис. 296) во вторую среду прямые по всевозможным направлениям и на них откладывать длины волнового вектора $\mathbf{k}$. Поскольку направления $\mathbf{k}$ и $\mathbf{N}$ совпадают, а каждому направлению $\mathbf{N}$ соответствуют два значе-

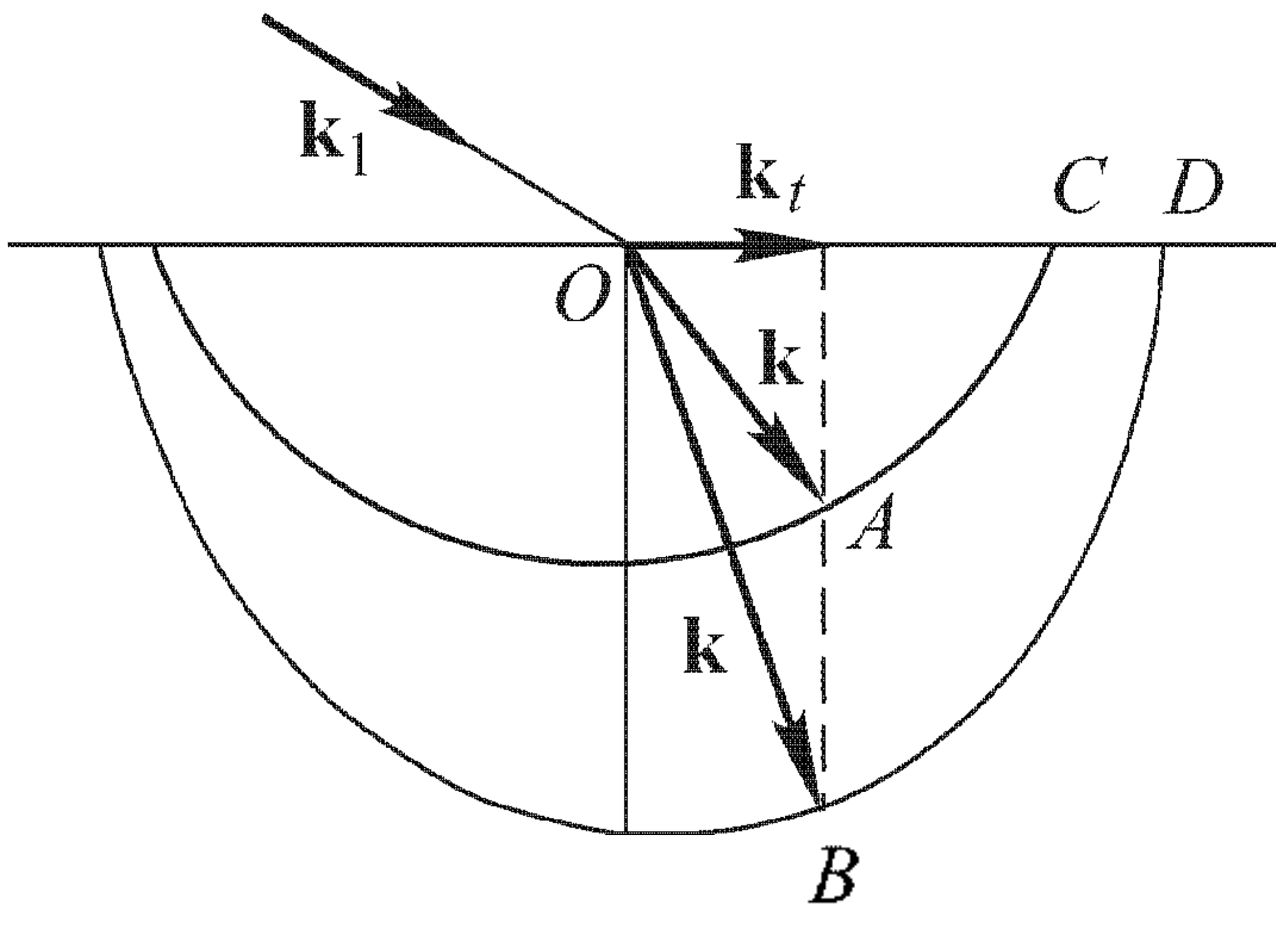


Рис. 296

ния нормальной скорости, можно сказать также, что каждому направлению нормали **N** соответствуют два значения волнового числа k. Поэтому в результате построения получится сложная поверхность, состоящая из двух слоев. Будем называть ее поверхностью *волновых векторов*[1]. Она просто связана с поверхностью нормалей. Поскольку $k = \omega/v$, поверхность волновых векторов по отношению к поверхности нормалей является *обратной*: длины ее радиусов-векторов обратно пропорциональны длинам параллельных им радиусов-векторов поверхности нормалей.

Пересечем поверхность волновых векторов плоскостью падения — в сечении получатся две кривые (рис. 296). Отложим от точки O вдоль границы раздела тангенциальную составляющую $\mathbf{k}_t = \mathbf{k}_{1t}$ волнового вектора $\mathbf{k}$ и из конца этой составляющей восстановим перпендикуляр к границе раздела. Он пересечет указанные две кривые, вообще говоря, в двух точках A и B. Соединив эти точки с точкой O, получим два вектора $\overrightarrow{OA}$ и $\overrightarrow{OB}$, каждый из которых может быть волновым вектором преломленной волны.

Тем самым определятся направления волновых нормалей **N** и соответствующие им значения нормальных скоростей. Поверхность волновых векторов позволяет построить поверхность нормалей, а затем лучевую поверхность и найти направления соответствующих лучей и лучевые скорости. Векторы **N** и **s** определят направление магнитного поля **H**, поскольку оно перпендикулярно к плоскости $(\mathbf{N}, \mathbf{s})$. Определятся и направления векторов **D** и **E**, т.е. поляризация обеих волн в кристалле. В общем случае направления лучей и волновых нормалей не совпадают. В оптически двуосных кристаллах оба луча, как правило, выходят из плоскости падения.

Если волна падает на кристалл из изотропной среды, то она может быть поляризована как угодно. В кристалл она вступит в виде двух линейно поляризованных волн. Если же первая среда также кристаллическая, то в направлении $\mathbf{k}_1$ могут распространяться две волны с различными поляризациями и с различными нормальными скоростями. Их надо рассматривать как *независимые волны*. Каждая из них в кристалле расщепится на *две линейно поляризованные волны*. Во второй среде получится всего *четыре волны*.

---

[1] Если вместо величины $k = \omega/v$ откладывать величину c/v, т.е. показатель преломления n, то получится поверхность, подобная поверхности волновых векторов. В кристаллооптике ее называют *индикатрисой* или *поверхностью показателей преломления*. Радиус-вектор этой поверхности дает значение показателя преломления для плоской волны, нормаль которой совпадает по направлению с этим радиусом-вектором.

Может случиться, что длина вектора $\mathbf{k}_{1t}$ окажется больше длины одного или обоих отрезков OC и OD, отсекаемых на оси X поверхностью волновых векторов. Тогда построение, выполненное на рис. 296, станет невозможным. Возникнет либо одна преломленная волна, либо ни одной. Одна из волн или обе волны испытают *полное отражение.*

3. Таким же способом может быть решен вопрос о направлении и поляризации отраженного света. Надо только построить поверхность волновых векторов не во второй, а в первой среде. Если первая среда изотропна, то поверхность волновых векторов будет сферой. Возникнет только одна отраженная волна, причем угол падения будет равен углу отражения. Если же первая среда анизотропна, то при отражении волна расщепится на две линейно поляризованные волны, идущие в разных направлениях.

ГЛАВА VIII

МОЛЕКУЛЯРНАЯ ОПТИКА

§ 84. Классическая теория дисперсии света

1. Все среды, за исключением абсолютного вакуума, обладают *дисперсией*, т.е. зависимостью показателя преломления электромагнитных волн от их частоты. Наилучшим приближением к вакууму является межпланетное и межзвездное пространство. По астрофизическим данным средняя плотность вещества в межпланетном пространстве составляет около одного атома (иона) на 1 см^3. В межзвездном пространстве нашей Галактики средняя концентрация вещества — около 10^{-2} атомов на 1 см^3, а в межгалактическом пространстве еще меньше. В лучших же вакуумных приборах она не меньше 10^4 атомов на 1 см^3, а обычно гораздо больше.

Отсутствие дисперсии у межзвездного пространства доказывается астрономическими наблюдениями над затмениями *двойных звезд*. Допустим, например, что красные лучи в межзвездном пространстве распространяются быстрее синих. Тогда при начале затмения должно было бы наблюдаться изменение цвета звезды от нормального к синему, а при окончании — от красного к нормальному. При колоссальных расстояниях до звезд этот эффект не мог бы ускользнуть от наблюдения, даже если бы разница в скоростях красных и синих лучей оказалась ничтожно малой. В действительности в видимой области спектра он обнаружен не был. Еще Араго на основании своих наблюдений над двойной звездой Альголь пришел к заключению, что разность между скоростями красных и синих лучей во всяком случае не может превышать одной стотысячной скорости света.

Конечно, отсутствие дисперсии видимого света в межзвездном пространстве экспериментально доказано лишь с той точностью, которую может обеспечить эксперимент. Наблюдения над пульсарами с помощью современных радиоприемных устройств позволили установить наличие дисперсии межзвездного пространства в области радиодиапазона (см. § 87).

2. Феноменологическая теория Максвелла, не содержащая никаких атомно-молекулярных констант, не давала объяснения дисперсии света. Для этого необходимы атомистические представления. Дисперсия света возникает в результате *вынужденных колебаний* заряженных частиц — электронов и ионов — под действием переменного поля электромагнитной волны.

В классической теории Г.А. Лорентца строение колеблющихся систем — атомов и молекул — и их колебания описываются на основе классических представлений о движении и законов Ньютона. В нашем курсе мы можем в основном ограничиться только такой классической теорией. Теоретическому рассмотрению проще всего поддается *дисперсия в газах*, так как в этом случае в первом приближении можно не учитывать сложное взаимодействие атомов и молекул среды. Для не очень плотных газов основные предположения теории выполняются с меньшей натяжкой, чем в случае конденсированных сред. Поэтому экспериментальную проверку этих предположений лучше всего производить именно на газах, для которых и теория разработана лучше. В дальнейшем мы в основном ограничимся этим простейшим случаем.

Все электроны, входящие в атом, можно разделить на *периферийные*, или *оптические*, и *электроны внутренних оболочек*. На излучение и поглощение света в оптической области спектра оказывают влияние практически *одни только оптические электроны*. Собственные частоты электронов внутренних оболочек слишком велики, так что их колебания в поле световой волны практически не возбуждаются. Поэтому в теории дисперсии можно ограничиться рассмотрением одних только оптических электронов.

Для простоты предположим сначала, что в атоме есть *всего один оптический электрон*. В классической теории дисперсии оптический электрон рассматривается как *затухающий гармонический осциллятор*, колебания которого в поле световой волны описываются уравнением

$$m\ddot{\mathbf{r}} = -k\mathbf{r} - g\dot{\mathbf{r}} + e\mathbf{E}',$$

где m — масса, e — заряд электрона, $-k\mathbf{r}$ — *квазиупругая возвращающая сила*, стремящаяся вернуть электрон в положение равновесия, $-g\dot{\mathbf{r}}$ — сила, аналогичная *силе трения* и введенная для учета *поглощения света*, $\mathbf{E}'$ — напряженность электрического поля, *действующего* на электрон. Поле $\mathbf{E}'$, вообще говоря, отличается от среднего макроскопического поля $\mathbf{E}$, входящего в уравнения Максвелла (см. т. III, § 35), но в случае неплотных газов этим различием можно пренебречь. Разделив на m, приведем предыдущее уравнение к виду

$$\ddot{\mathbf{r}} + 2\gamma\dot{\mathbf{r}} + \omega_0^2\mathbf{r} = \frac{e}{m}\mathbf{E}', \tag{84.1}$$

где $\omega_0^2 = k/m$, $2\gamma = g/m$. Магнитной силой $\frac{e}{c}[\mathbf{vH}]$, действующей на электрон, мы пренебрегли, так как скорость электрона v пренебрежимо мала по сравнению со скоростью света c. Эта сила проявляется лишь в *световом давлении*, которое в обычных условиях мало и в разбираемом нами вопросе не играет существенной роли (см. задачу 2 к этому параграфу).

Все силы, действующие внутри атомов и молекул, имеют *электрическую природу*. Такими силами определяется и внутренняя структура этих частиц. Однако объяснить этими силами существование и структуру атомов и молекул на основе классических представлений оказалось невозможным. Это было сделано только в рамках *квантовой физики*. Никаких квазиупругих сил и сил трения, пропорциональных скорости заряженных частиц, в атомах и молекулах нет. Правильная теория дисперсии должна принимать во внимание только *реально существующие силы* и основываться на *квантовых законах*. Такую теорию дает квантовая механика. Однако она приводит к поразительному результату, что *в отношении дисперсии и поглощения света атомы и молекулы ведут себя так, как если бы среда представляла собой набор осцилляторов с различными собственными частотами и коэффициентами затухания, подчиняющихся классическим уравнениям движения Ньютона.*

Собственные частоты и коэффициенты затухания не могут быть вычислены на основе классической модели. В классической теории на них надо смотреть как на *формально введенные постоянные*. Вычисление этих постоянных и раскрытие их истинного физического смысла возможно только в *рамках квантовой теории*. На классическую же теорию надо смотреть как на теорию, дающую *модель диспергирующей среды*, которая приводит к правильным окончательным результатам, если к ней применять законы классической механики.

После этого замечания можно перейти к изложению классической теории дисперсии, не опасаясь, что смысл этой теории будет понят неправильно.

3. Предположим, что поле $\mathbf{E}$ представляется плоской волной $\mathbf{E} = \mathbf{A}e^{i(\omega t - \mathbf{kr})} = \mathbf{E}_0(\mathbf{r})e^{i\omega t}$. Амплитуда поля $\mathbf{E}_0 = \mathbf{A}e^{-i\mathbf{kr}}$ меняется от точки к точке. Значит, в различных точках траектории электрон подвергается действию поля различной амплитуды. (Указанное обстоятельство приводит к явлению так называемой *пространственной дисперсии*.) Однако мы пренебрежем этим обстоятельством, предполагая, что амплитуда колебаний электрона $\mathbf{r}_0$ мала по сравнению с длиной волны λ. Для теории дисперсии имеет значение не общее, а только *частное решение* уравнения (84.1), представляющее *вынужденные колебания* осциллятора $\mathbf{r} = \mathbf{r}_0 e^{i\omega t}$. Амплитуда $\mathbf{r}_0$ найдется подстановкой этого выражения в (84.1). Если пренебречь различием между $\mathbf{E}$ и $\mathbf{E}'$, то получится

$$\mathbf{r} = \frac{e/m}{\omega_0^2 - \omega^2 + 2i\omega\gamma}\mathbf{E}. \tag{84.2}$$

Атом в электрическом поле приобретает дипольный момент $\mathbf{p} = e\mathbf{r} = \beta\mathbf{E}$, где β — поляризуемость атома, определяемая формулой

$$\beta = \frac{e^2/m}{\omega_0^2 - \omega^2 + 2i\omega\gamma}. \tag{84.3}$$

Если N — число атомов в единице объема, то вектор поляризации среды будет $\mathbf{P} = N\mathbf{p} = N\beta\mathbf{E}$, а индукция

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \varepsilon\mathbf{E}, \tag{84.4}$$

где

$$\varepsilon = 1 + \frac{4\pi N e^2/m}{\omega_0^2 - \omega^2 + 2i\omega\gamma}. \tag{84.5}$$

Материальное уравнение $\mathbf{D} = \varepsilon\mathbf{E}$ принимает такой же вид, как и в формальной теории Максвелла, с тем существенным отличием, что диэлектрическая проницаемость ε теперь *зависит от частоты* ω. Поэтому сохраняют силу все ранее полученные результаты, если в них постоянную ε заменить функцией $\varepsilon(\omega)$. Функция $\varepsilon(\omega)$ получилась *комплексной*. Этого и следовало ожидать, так как в нашей модели учтено *поглощение света*. Введем комплексный показатель преломления по формуле

$$\sqrt{\varepsilon} = n - i\varkappa, \tag{84.6}$$

где n — *вещественный показатель преломления*, а $\varkappa$ — *показатель затухания среды*. Обе функции $n(\omega)$ и $\varkappa(\omega)$ зависят от частоты. Таким образом, получают принципиальное, хотя и формальное объяснение не только дисперсия, но и абсорбция света.

Возведем (84.6) в квадрат и сравним вещественные и мнимые части полученного соотношения и соотношения (84.5). Тогда получим

$$n^2 - \varkappa^2 = 1 + 4\pi\frac{Ne^2/m}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\gamma^2}(\omega_0^2 - \omega^2), \tag{84.7}$$

$$n\varkappa = 4\pi\frac{Ne^2/m}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\gamma^2}\omega\gamma. \tag{84.8}$$

С помощью этих формул показатели преломления и затухания n и $\varkappa$ можно выразить через частоту ω и молекулярные параметры N и e^2/m.

Вдали от собственной частоты ω_0, где $2\omega\gamma \ll |\omega_0^2 - \omega^2|$, мнимой частью в (84.5) можно пренебречь и получить приближенную формулу:

$$\varepsilon = n^2 = 1 + \frac{4\pi N e^2/m}{\omega_0^2 - \omega^2}. \tag{84.9}$$

Формула для n такого вида впервые была получена Зельмейером в 1871 г. 1) в механической теории эфира. Конечно, в формуле самого Зельмейера нет никакого намека на связь показателя преломления с диэлектрической проницаемостью ε и вообще со всякими величинами, характеризующими электрические свойства среды.

На рис. 297 сплошной линией представлен график функции $\varepsilon = \varepsilon(\omega)$. При переходе через точку $\omega = \omega_0$ эта функция претерпевает скачок от $+\infty$ до $-\infty$. Показатель преломления n представлен штриховой, а показатель затухания $\varkappa$ — штрихпунктирной линией. Величина n обращается в нуль на участке, где $\varepsilon(\omega) < 0$. Показатель затухания $\varkappa$, наоборот, на этом участке отличен от нуля, обращаясь в нуль при всех остальных частотах. Величина $\varkappa$ описывает именно *затухание, а не поглощение волн*, так как кривые рис. 297 относятся к модели *непоглощающей среды* (диэлектрическая проницаемость ε всюду вещественна).

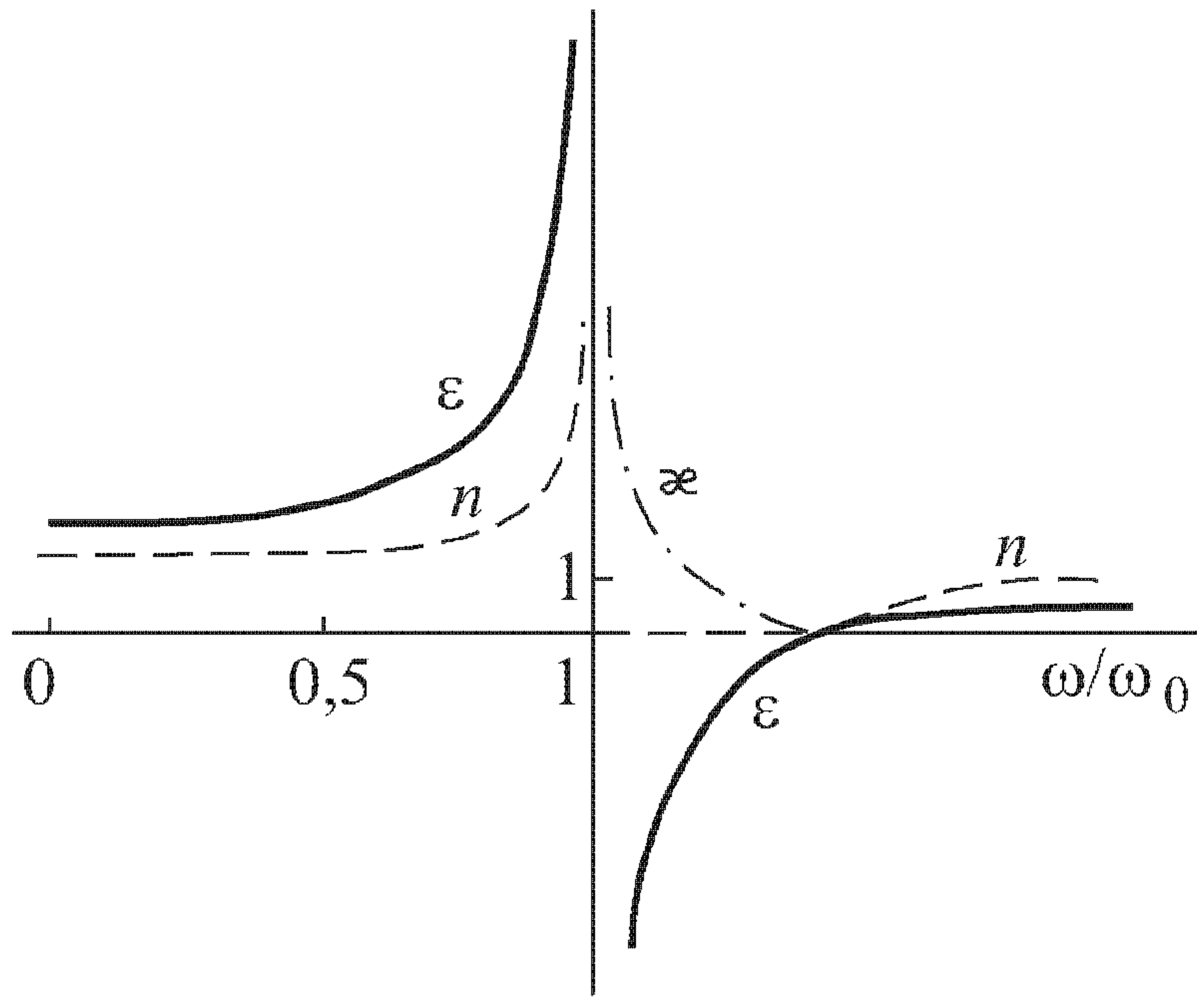


Рис. 297

Впрочем, эти выводы имеют формальный характер, так как вблизи собственной частоты ω_0 соотношение $2\omega\gamma \ll |\omega_0^2 - \omega^2|$ не выполняется, а потому формула (84.9) неприменима. В действительности при переходе через точку $\omega = \omega_0$ величины ε, n, $\varkappa$ должны меняться *непрерывно*, нигде не обращаясь в бесконечность. В частности, поднимающаяся ветвь кривой $\varepsilon = \varepsilon(\omega)$ левее точки $\omega = \omega_0$ должна переходить в опускающуюся, чтобы сомкнуться с другой поднимающейся ветвью, расположенной правее той же точки.

---

1) Как отметил Рэлей, формула (84.9) была получена двумя годами раньше Максвеллом, который еще в 1869 г. изложил основы механической теории дисперсии в виде экзаменационной задачи. Формулы Максвелла содержали также члены с затуханием.

Это замечание позволяет без особого труда представить ход кривых n и $\varkappa$ в окрестности полосы поглощения. Типичный вид кривых $n = n(\omega)$ и $\varkappa = \varkappa(\omega)$ приведен на рис. 298. Там, где поглощение невелико, показатель преломления $n(\omega)$ возрастает

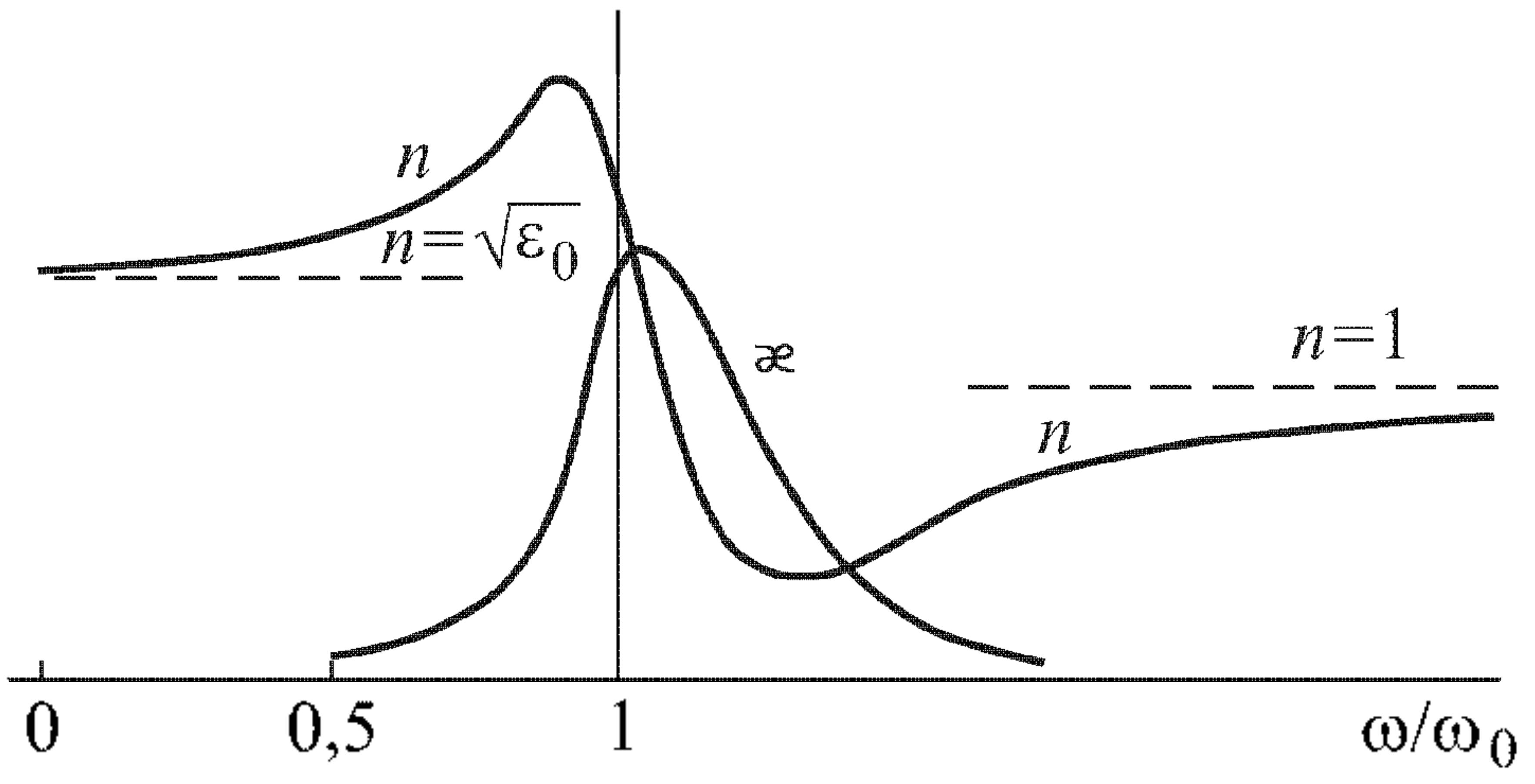


Рис. 298

с частотой. В этом случае говорят, что *дисперсия нормальная.* В области сильного поглощения $n(\omega)$ уменьшается с частотой. Такую дисперсию называют *аномальной.* Ее трудно наблюдать из-за сильного поглощения.

4. Теоретические соображения, изложенные выше, справедливы не только для электронов, *но и для ионов*, и притом в последнем случае классические представления более обоснованы ввиду относительно больших масс ионов. Во всех телах наблюдается не одна, а несколько полос поглощения. Чтобы это учесть, в классической модельной теории принимается, что вещество построено из частиц *различного типа* — электронов и ионов, которые ведут себя как затухающие гармонические осцилляторы с *различными собственными частотами.* В газах можно пренебречь их взаимодействием. Тогда формула (84.5) принимает вид

$$\varepsilon = 1 + \sum_k \frac{4\pi N_k e_k^2/m_k}{\omega_k^2 - \omega^2 + 2i\omega\gamma_k}, \tag{84.10}$$

где N_k, m_k, e_k, ω_k, γ_k — концентрация, масса, заряд, собственная частота и коэффициент затухания осциллятора k-го типа, а суммирование ведется по всем k. Каждой собственной частоте соответствует своя линия поглощения, вблизи которой показатель преломления меняется аномально. Общий ход показателя преломления в зависимости от частоты представлен схематически на рис. 299.

Сравнивая формулу (84.10) с опытом, можно найти коэффициенты $C_k \equiv N_k e_k^2/m_k$. По этим данным можно оценить удельные заряды e_k/m_k и вывести заключения относительно природы осцилляторов. Действительно, заряды e_k по порядку величины не могут существенно отличаться от элементарного заряда $e = 4{,}8 \cdot 10^{-10}$ СГСЭ, а N_k — от числа атомов в единице объема, т.е. от $N_{\text{Ав}}\rho/A$, где $N_{\text{Ав}}$ — число Авогадро, ρ — плотность вещества, A — атомная масса (по старой терминологии — атомный вес). Таким образом,

$$C_k = \frac{N_k e_k^2}{m_k} \approx \rho \frac{N_{\text{Ав}} e}{A} \frac{e_k}{m_k} = \rho \frac{F}{A} \frac{e_k}{m_k},$$

где $F \equiv N_{\text{Ав}} e = 2{,}9 \cdot 10^{14}$ СГС·г-экв$^{-1}$ — число Фарадея. С помощью этой формулы и можно оценить величины e_k/m_k. Таким путем, в согласии с теоретическими представлениями, было найдено, что все осцилляторы отчетливо подразделяются на две группы: у одной удельные заряды по порядку величины такие же, как у электронов, у другой — как у ионов. Первым соответствуют полосы поглощения, лежащие в *ультрафиолетовой* (реже — *в видимой*), вторым — в *инфракрасной области спектра.*

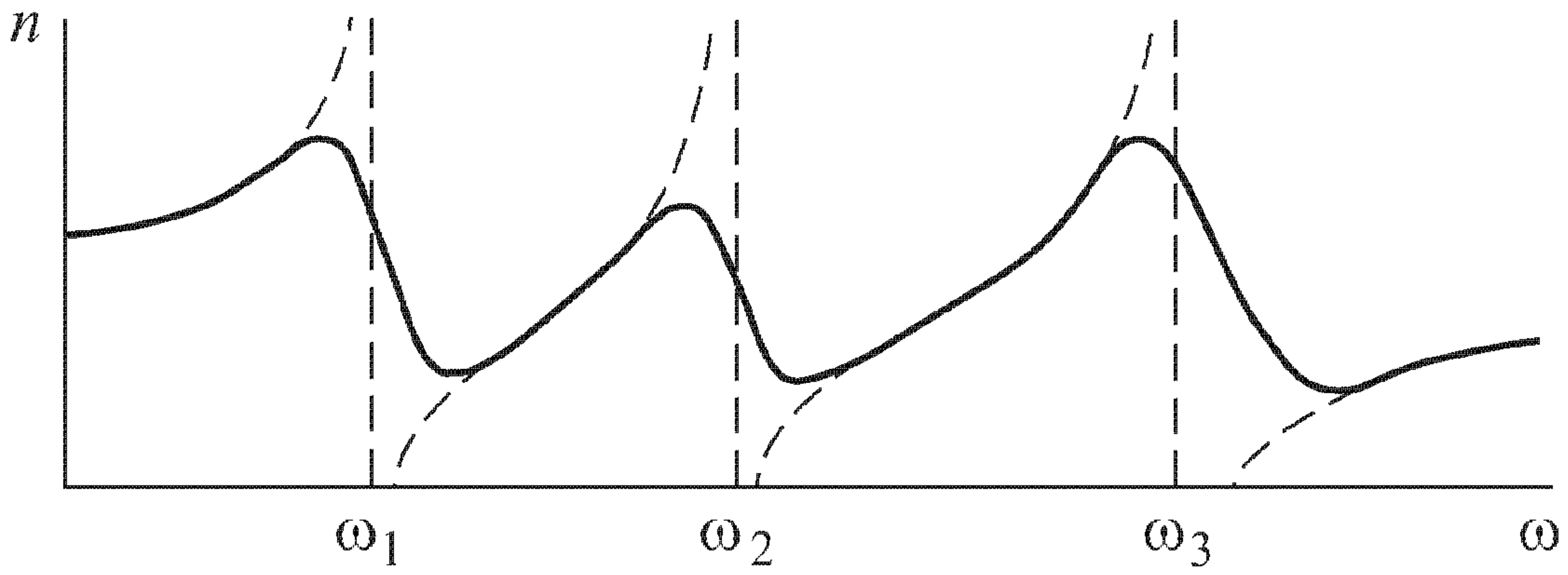


Рис. 299

5. Формула (84.10) и соответствующие ей формулы для n и $\varkappa$ применимы и в случае *металлов.* В инфракрасной области спектра основную роль играют *свободные электроны,* для которых следует положить $\omega_k = 0$. Влияют также ионы вблизи их собственных частот. В видимой и примыкающей к ней ультрафиолетовой областях спектра формула (84.10) удовлетворительно согласуется с опытом, если наряду со свободными учесть также электроны, *связанные в ионах металла* (колебания самих ионов в этих областях спектра не играют роли).

6. Учтем теперь, хотя далеко не безупречным способом, взаимодействие молекул, поляризованных электрическим полем световой волны. Для этого воспользуемся методом *действующего поля.* В кристаллах кубической системы, построенных из точечных и изотропных атомов, поле $\mathbf{E}'$, действующее на молекулу, связано со средним максвелловским полем $\mathbf{E}$ соотношением

$$\mathbf{E}' = \mathbf{E} + \frac{4\pi}{3}\mathbf{P} \tag{84.11}$$

(см. т. III, § 35). То же соотношение приближенно справедливо для изотропных газов и жидкостей, построенных из изотропных молекул, если под $\mathbf{E}'$ понимать действующее поле, усредненное по положениям центров молекул в физически бесконечно малом объеме среды. Формула (84.11) выведена в электростатике, но ею можно приближенно пользоваться и в случае переменных полей, когда длина волны велика по сравнению с межмолекулярными расстояниями.

Будем предполагать, что все молекулы одинаковы и каждая из них содержит только один оптический электрон. Кроме того, пренебрежем затуханием, предполагая, что частота ω далека от полосы поглощения. В этих предположениях все, что нужно сделать, — это заменить в уравнении (84.1) $\mathbf{E}'$ выражением (84.11). Последующие вычисления крайне элементарны, но мы опустим их, так как они уже были выполнены в электростатике (см. т. III, § 35). Там по методу, в точности совпадающему с излагаемым здесь, была получена формула Мосотти–Клаузиуса:

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{4\pi}{3} N\beta. \tag{84.12}$$

Подставив сюда $\varepsilon = n^2$ и выражение для β из формулы (84.3), придем к соотношению

$$\frac{n^2 - 1}{n^2 + 2} = \frac{4\pi}{3} \frac{Ne^2/m}{\omega_0^2 - \omega^2}. \tag{84.13}$$

Эта формула была одновременно (1880 г.) получена в электронной теории голландским физиком Г.А. Лорентцом и датским физиком Л. Лоренцом (1829–1891) еще из домаксвелловских представлений о свете. Она называется *формулой Лорентца–Лоренца.* Из нее следует, что для данного вещества при неизменной частоте ω должно выполняться соотношение

$$r \equiv \frac{1}{\rho} \frac{n^2 - 1}{n^2 + 2} = \mathrm{const}, \tag{84.14}$$

поскольку концентрация N молекул пропорциональна плотности вещества ρ. Величина, стоящая в левой части этого соотношения, называется *удельной рефракцией.* Таким образом, согласно формуле Лорентца–Лоренца *удельная рефракция вещества не должна изменяться при изменении плотности.*

В ряде случаев формула (84.14) хорошо выполняется. Так, для воздуха при изменении давления от 1 до 200 атм постоянство удельной рефракции выполняется с ошибкой, не превышающей 0,1 %. Нередко удельная рефракция остается приблизительно постоянной даже при переходе вещества из парообразного состояния в жидкое. Так, при ожижении газообразного кислорода (изменение плотности в 800 раз) или конденсации водяного пара (изменение плотности в 1200 раз) удельная рефракция остается постоянной с точностью до 2–3 %. Эти факты удивительны, так как в приведенных примерах предположения, в которых выведена формула (84.11), несомненно, не выполняются. Имеется, однако, много случаев отступлений от формулы (84.14).

7. Еще более удивительно эмпирическое правило, по которому можно приближенно вычислять рефракцию смеси или химического соединения. Если вещество состоит из одинаковых частиц (атомов или молекул), то $\delta = Nm$, где m — масса одной частицы. Поэтому из формулы (84.12) с учетом соотношения $\varepsilon = n^2$ получаем

$$\beta = \frac{3}{4\pi} mr. \tag{84.15}$$

Поляризуемость частицы β, а с ней и величина mr есть свойство *только самой частицы*. Правило, которое было упомянуто выше, состоит в том, что *поляризуемость частицы аддитивна, т.е. поляризуемость сложной частицы равна сумме поляризуемостей частиц, из которых она состоит.* Следовательно,

$$mr = \sum m_i r_i, \tag{84.16}$$

где m и m_i — массы составной и составляющих частиц, а r и r_i — удельные рефракции соответствующих им химических веществ. Удивительно, что правило это относится не только к механическим смесям, но и *к химическим соединениям.* Возьмем в качестве составляющих частиц атомы, а составных — молекулы. Тогда $m_i = A_i m_{\mathrm{H}}$, $m = M m_{\mathrm{H}}$, где m_{H} — масса атома водорода, а M и A_i — соответствующие молекулярная и атомные массы. Поэтому формула (84.16) переходит в

$$Mr = \sum A_i r_i. \tag{84.17}$$

Произведение удельной рефракции на атомную массу называется *атомной*, а на молекулярную массу — *молекулярной рефракцией.* Таким образом, *молекулярная рефракция химического соединения равна сумме атомных рефракций.*

В книге Борна[1]) приведены обширные таблицы, подтверждающие формулу Лоренца–Лоренца и аддитивность атомной рефракции. Там же приведены примеры и значительных отступлений от этих правил.

8. Помимо дисперсии, связанной с вынужденными колебаниями электронов и ионов, существует другой вид дисперсии, на который обратил внимание Дебай. Она проявляется в средах с *полярными молекулами*, главным образом в жидкостях. Полярные — это такие молекулы, которые обладают «твердыми» дипольными моментами, обусловленными несимметричным распределением зарядов внутри молекулы. Полярные молекулы могут обладать и «индуцированными» дипольными моментами, возбуждаемыми внешним электрическим полем. Как правило, индуцированные моменты полярных молекул малы по сравнению с твердыми.

---

[1]) *Макс Борн.* Оптика.— Харьков; Киев: Гос. научно-техн. изд-во Украины. 1937.

В отсутствие электрического поля дипольные моменты полярных молекул изотропной среды ориентируются в пространстве по всем направлениям и притом хаотически. В этом случае среда никакой электрической поляризацией не обладает. При наложении электрического поля оно стремится ориентировать дипольные моменты молекул вдоль поля, причем эта ориентация постоянно нарушается из-за теплового движения. В результате в среде возникает электрическая поляризация (см. т. III, § 36). Если электрическое поле меняется во времени с частотой ω, то с той же частотой будет меняться эта «ориентационная» поляризация и обусловленная ею диэлектрическая проницаемость среды. Амплитуды вынужденных вращений полярных молекул зависят от частоты ω приложенного электрического поля. С этим связана дисперсия *вращательной части* диэлектрической проницаемости и показателя преломления.

Не входя в подробное количественное рассмотрение, ограничимся качественными замечаниями.

В полях низких частот успевает установиться такая же мгновенная поляризация среды, как и в статических полях. В такой области частот величины ε и n практически постоянны, т.е. не зависят от ω. Если дипольные моменты полярных молекул велики, как, например, у воды и спиртов, то в этой области частот значения ε и n велики. Так, в случае воды $\varepsilon = 81$, $n = \sqrt{\varepsilon} = 9$.

В другом предельном случае очень высоких частот дипольные молекулы не успевают заметно поворачиваться за времена порядка периода световых колебаний. В этой области частот практически нет вынужденных вращений молекул, т.е. вращательные части ε и n выпадают. Поэтому должна существовать промежуточная область частот, в которой величины ε и n аномально сильно уменьшаются до тех значений, которые они принимают в оптической области спектра. Такая область частот у воды и спиртов лежит в диапазоне сантиметровых радиоволн. За пределами этой области (со стороны высоких частот) молекулы начинают вести себя так, как если бы они не были полярными. Этим объясняется резкое расхождение между показателем преломления воды n в оптической области спектра и значением $\sqrt{\varepsilon}$ в электростатике и области низкочастотных электромагнитных волн.

Этим же объясняется, почему в оптике магнитную проницаемость μ почти всегда можно считать равной единице для всех веществ. Действительно, механизм намагничивания — такой же, как и поляризации полярных диэлектриков. Он состоит в ориентации магнитных моментов атомов и молекул во внешнем магнитном поле. Но этот эффект не имеет места в высокочастотных полях, видимой и примыкающих к ней областях спектра.

9. Подводя итоги, дадим схематический обзор дисперсии во всем диапазоне частот электромагнитных волн. В области радиоволн длиннее примерно 1 см существенна только одна собственная частота $\omega_0 = 0$, которой обладают свободные электроны или ионы. В указанном радиодиапазоне дисперсией практически могут обладать *только ионизованные газы* (см. § 86). Если свободных электронов или ионов нет, то в этом диапазоне нет и дисперсии.

В полярных диэлектриках дисперсия появляется примерно с сантиметрового диапазона радиоволн. Показатель преломления убывает с частотой, т.е. дисперсия *аномальна*, а поглощение велико.

По мере приближения к собственным частотам ионов дисперсия становится *нормальной*, а внутри самой полосы поглощения, обусловленной колебаниями ионов, — *аномальной*. После прохождения через инфракрасную область поглощения вдали от нее вынужденные колебания ионов практически прекращаются. Здесь, а также в видимой и ультрафиолетовой областях спектра дисперсия вызывается колебаниями только *связанных оптических электронов*. В прозрачной области (точнее, области слабого поглощения) она имеет *нормальный*, а в области сильного поглощения — *аномальный* характер. Полоса поглощения лежит обычно в области ультрафиолета, а иногда в видимой области.

При переходе через полосу поглощения, т.е. область аномальной дисперсии, сбусловленную колебаниями электронов, дисперсия снова становится *нормальной*. Если частота ω велика по сравнению со всеми собственными частотами электронов, то в формуле (84.10) величинами ω_k^2 можно пренебречь по сравнению с ω^2. Мы вступили в область *мягкого рентгеновского излучения*. Если не учитывать поглощение и заметить, что в колебаниях принимают участие одни только электроны, то дисперсионной формуле (84.10) в указанной области спектра можно придать вид

$$n^2 = 1 - \frac{Ne^2}{m\omega^2}. \tag{84.18}$$

Показатель преломления здесь *меньше единицы*, хотя и очень мало отличается от нее. Его удалось измерить, наблюдая отклонение рентгеновских лучей в призме из различных материалов, хотя этот метод и не может дать большой точности.

Более падежные результаты дают измерения предельного угла полного отражения, испытываемого рентгеновскими лучами при переходе из воздуха в твердое тело ($n < 1$). Оказалось, что показатель преломления рентгеновских лучей меньше единицы на величину порядка $(10^{-5} - 10^{-6})\rho$, где ρ — плотность вещества. Так, в случае стекла (крон с плотностью 2,52) для $\lambda = 0{,}1279$ нм он оказался равным $n = 0{,}999995 = 1 - 5 \cdot 10^{-6}$.

При увеличении жесткости рентгеновского излучения резонансные явления могут проявляться на электронах *внутренних оболочек атомов*, связанных с ними более сильно, чем оптические электроны. Действительно, удалось наблюдать аномальную дисперсию рентгеновских лучей вблизи частот их истинного фотоэлектрического поглощения.

ЗАДАЧИ

1. Найти максимальную скорость вынужденных колебаний свободного электрона в поле солнечного излучения вблизи земной поверхности (см. задачу к § 5). Определить также отношение максимальной силы F_B, действующей на такой электрон со стороны магнитного поля, к максимальной силе F_E, действующей со стороны электрического поля. Поле солнечного излучения заменить монохроматическим $\mathbf{E} = \mathbf{E}_0 \cos \omega t$ с длиной волны $\lambda = 550$ нм.

Решение. В уравнении движения свободного электрона $m\dot{\mathbf{v}} = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{vB}]\right)$ пренебрежем действием магнитного поля. Тогда

$$m\dot{\mathbf{v}} = e\mathbf{E} = e\mathbf{E}_0 \cos \omega t, \quad \mathbf{v} = \frac{e}{m\omega}\mathbf{E}_0 \sin \omega t, \quad v_{\text{макс}} = \frac{eE_0}{m\omega} = \frac{eE_0\lambda}{2\pi mc}.$$

Так как $\overline{E^2} = \frac{1}{2}E_0^2 = 5{,}85 \cdot 10^{-4}$, то $E_0 = \sqrt{2\overline{E^2}} = 0{,}034$ СГСЭ. Подставляя это значение, получим

$$v_{\text{макс}} = 5{,}3 \text{ см/с}, \quad \frac{F_B}{F_E} = \frac{v_{\text{макс}}}{2c} = 0{,}9 \cdot 10^{-10}.$$

2. Определить среднюю силу светового давления на гармонический затухающий осциллятор, колеблющийся в поле монохроматической однородной световой волны. Рассчитать также среднюю энергию ε, поглощаемую осциллятором в одну секунду, и выразить через нее среднюю силу светового давления. Показать, что если бы поглощения света не было, то средняя сила светового давления равнялась бы нулю.

Решение. В месте нахождения осциллятора электромагнитное поле имеет вид $\mathbf{E} = A\mathbf{e}e^{i\omega t}$, $\mathbf{B} = A\mathbf{b}e^{i\omega t}$, где $\mathbf{e}$ и $\mathbf{b}$ — взаимно перпендикулярные единичные векторы. Пренебрегая сначала действием магнитного поля, дифференцированием (84.2) находим

$$\mathbf{v} = \frac{i\omega e/m}{(\omega_0^2 - \omega^2) + 2i\omega\gamma}\mathbf{E}.$$

Сила, действующая на электрон со стороны магнитного поля, $\mathbf{F} = \frac{e}{c}\,[\mathbf{vB}]$. Поскольку, однако, вычисления ведутся в комплексной форме, следует заменить это выражение на

$$\mathbf{F} = \frac{e}{4c}\,[(\mathbf{v}+\mathbf{v}^*)(\mathbf{B}+\mathbf{B}^*)].$$

Для средней силы получаем

$$\overline{\mathbf{F}} = \frac{e}{4c}\,\{[\mathbf{v}^*\mathbf{B}] + [\mathbf{vB}^*]\},$$

или после простых преобразований

$$\overline{\mathbf{F}} = \frac{m\gamma}{c}\,\mathbf{vv}^*[\mathbf{eb}].$$

Отсюда видно, что средняя сила светового давления направлена в сторону векторного произведения $[\mathbf{eb}]$, т.е. в сторону распространения волны.

Энергия, поглощаемая в единицу времени:

$$\int\limits_0^1 2m\gamma\left(\frac{\mathbf{v}+\mathbf{v}^*}{2}\right)^2 dt,$$

а ее среднее значение

$$\varepsilon = \int\limits_0^1 m\gamma\mathbf{vv}^*\,dt = m\gamma\mathbf{vv}^*.$$

Таким образом,

$$\overline{F} = \varepsilon/c. \tag{84.19}$$

Эта формула может быть получена из простых соображений. Излучение, поглощаемое атомом вещества, передает ему не только энергию ε, но и связанный с ней импульс $g = \varepsilon/c$, который проявляется в силах светового давления.

Если свет распространяется через поглощающий газ, то, согласно изложенному, он должен производить на него давление. Впервые такое предположение высказал Кеплер, изучая формы кометных хвостов. Экспериментальное доказательство существования светового давления на газы было дано П.Н. Лебедевым (1910 г.).

§ 85. Понятие о дисперсионной формуле квантовой механики

1. Классическая дисперсионная формула (84.10) для газов с большой точностью описывает фактически наблюдаемый ход показателя преломления вблизи отдельных линий поглощения ω_k, но лишь при том условии, если коэффициенты N_k, а также собственные частоты ω_k и коэффициенты затухания γ_k рассматриваются как *эмпирические постоянные, определяемые из самой кривой дисперсии и фактического положения спектральных линий в спектре излучения или поглощения вещества.* В частности, для согласования с опытом оказалось необходимым в этой формуле величины N_k заменить на $f_k N_k$, где f_k — постоянные коэффициенты, меньшие единицы, называемые *силами осцилляторов.*

Теорию дисперсии в квантовой механике можно строить по той же схеме, что и в классической физике. Задача сводится к вычислению *поляризуемости атомов и молекул* в электрическом поле световой волны. Но при решении этой задачи надо пользоваться не классическими, а *квантомеханическими* уравнениями движения, например, в форме *волнового уравнения Шредингера* (1887–1961). Поскольку в данной книге квантовая механика не предполагается известной, систематическое изложение квантовой теории дисперсии в ней невозможно. Можно дать только общую характеристику и некоторые результаты этой теории.

2. Квантовая теория дисперсии (и в этом одно из преимуществ ее перед классической) не пользуется моделью квазиупруго связанного электрона и не вводит никаких в действительности не существующих сил. Квантовая механика объясняет строение и устойчивость атомных систем с помощью *одних только электрических сил.* Для простоты будем предполагать, что в атоме есть всего только один *валентный электрон*, определяющий его химические и оптические свойства. Остальные электроны прочно удерживаются на внутренних оболочках, обладая значительно бóльшими энергиями связи. Состояния таких электронов практически не возмущаются слабым электрическим полем световой волны. Их роль сводится только к изменению электрического поля атомного ядра, в котором движется валентный электрон. Именно по этой причине атом может рассматриваться как *одноэлектронный.* Такая модель атома весьма близко соответствует действительности.

В отсутствие внешних полей атом может находиться только во вполне определенных *стационарных состояниях*, в которых его энергия может принимать также вполне определенные дискретные значения W_1, W_2, W_3, … (*уровни энергии*). Их принято нумеровать в порядке возрастания энергии. При переходе атома с высшего энергетического уровня k на более низкий уровень n происходит излучение света с испусканием кванта энергии

$$\hbar\omega_{kn} \equiv h\nu_{kn} = W_k - W_n. \tag{85.1}$$

При обратном переходе атома с низшего энергетического уровня n на более высокий уровень k происходит возбуждение атома с поглощением такого же кванта. Таким образом, в отличие от классического гармонического осциллятора, атом, даже если он одноэлектронный, излучает не одну частоту ω_0, а целый спектр частот ω_{kn}, которые в квантовой теории дисперсии и играют роль собственных частот атома. Если переход происходит с более низкого уровня k на более высокий уровень n (поглощение), то для сохранения без изменения соотношения (84.10) удобно ввести *отрицательные частоты* ω_{kn}. Если нет внешних возмущений (отсутствие силовых полей, невысокие температуры), то в результате процессов излучения все атомы перейдут на *низший* или *основной* энергетический уровень, т.е. в так называемое *основное* или *нормальное состояние.* На основном уровне изолированный атом будет находиться неограниченно долго, пока в результате внешнего воздействия он не перейдет на другой уровень.

При включении внешнего поля атом выйдет из исходного основного стационарного состояния. Возникшее состояние, согласно квантовой механике, можно характеризовать как суперпозицию стационарных состояний, каждому из которых приписывается определенная вероятность. Если электрическое поле слабое и гармонически меняется во времени с частотой ω, то эти вероятности также будут меняться во времени с той же частотой. Появится переменный дипольный момент атома, индуцированный электрическим полем световой волны. Его можно представить в виде $\mathbf{p} = \beta\mathbf{E}$. В общем случае поляризуемость атома β есть *тензор*, но мы ограничимся простейшим случаем, когда он *скаляр.* Если не учитывать поглощения (что допустимо вдали от собственных частот ω_{nk}), то квантовая механика приводит к результату

$$\beta = \frac{e^2}{m}\sum_k \frac{f_{nk}}{\omega_{nk}^2 - \omega^2}, \tag{85.2}$$

где f_{nk} — постоянные. Формула предполагает, что до включения электрического поля атом находился в стационарном состоянии n. Суммирование ведется по всем возможным значениям k, как бóльшим, так и меньшим n.

Для выяснения физического смысла постоянных f_{nk} заметим, что существует определенная вероятность, что за определенное время (например, равное единице) атом в отсутствие внешних полей *самопроизвольно* перейдет из высшего стационарного состояния k в более низкое стационарное состояние n. Постоянная f_{nk} пропорциональна этой вероятности. Самопроизвольный переход из низшего состояния n в высшее состояние k, конечно, невозможен. В этом случае f_{kn} определяется соотношением $f_{kn} = -f_{nk}$. Следовательно, при $n = k$ величины f_{nk} равны нулю, так что суммирование в (85.2) производится только по $k \neq n$. Величина f_{nk} и называется *силой осциллятора с частотой* ω_{nk}.

Используя выражения (85.2), приходим к следующей дисперсионной формуле для атомов с одним валентным электроном:

$$n^2 = 1 + 4\pi N\frac{e^2}{m}\sum_{k\neq n}\frac{f_{nk}}{\omega_{nk}^2-\omega^2}. \tag{85.3}$$

Сравнение этой формулы с классической формулой (84.10) (без учета поглощения) приводит к следующей наглядной аналогии. Все атомы, находящиеся в одном и том же исходном состоянии n, как бы разбиваются на группы, нумеруемые индексом k. Атомы одной и той же группы ведут себя как *классические гармонические осцилляторы* с собственной частотой ω_{nk}. Если $k > n$, то частота ω_{nk} считается *положительной*, а при $k < n$ — *отрицательной*. Число атомов k-й группы в единице объема определяется выражением $N_k = f_{nk}N$. Так как $\sum N_k = N$, то должно быть

$$\sum_k f_{nk} = 1. \tag{85.4}$$

Если в атоме Z валентных электронов, то это соотношение переходит в соотношение

$$\sum_k f_{nk} = Z. \tag{85.5}$$

Полученные формулы выражают так называемое *правило сумм*, найденное независимо друг от друга Томасом и Рейхе, с одной стороны, и Куном, с другой (1925 г.). Приведенные наводящие рассуждения не могут служить доказательством формул (85.4) и (85.5). Это видно уже из того, что при $k < n$ силы осцилляторов f_{nk}, а потому и числа атомов N_k отрицательны. Однако сами формулы (85.4) и (85.5) могут быть строго доказаны методами квантовой механики.

3. Если до включения электрического поля атомы газа не находились в одном и том же стационарном состоянии, а были распределены по таким состояниям, то формулу (85.3) следует заменить на формулу

$$n^2 = 1 + 4\pi N\frac{e^2}{m}\sum_n\sum_k\frac{f_{nk}}{\omega_{nk}^2-\omega^2}. \tag{85.6}$$

Таким же путем должны быть обобщены и правила сумм (85.4) и (85.5) — в них суммирование надо производить не только по всем значениям k, но и по всем значениям n.

Принципиально новое явление, предсказанное квантовой механикой, состоит в том, что в формулу (85.6) могут входить слагаемые не только с положительными f_{nk} (когда $k > n$), но и с отрицательными (когда $k < n$). Так как $\omega_{nk}^2 = \omega_{kn}^2$, то в дисперсионную формулу войдут только суммы $(f_{nk} + f_{kn})$, а не каждое слагаемое в отдельности. Если на нижнем энергетическом уровне находится больше атомов, чем на верхнем, то $(f_{nk} + f_{kn}) > 0$, в противоположном случае $(f_{nk} + f_{kn}) < 0$. В первом случае собственная частота $|\omega_{nk}|$ внесет в дисперсионную формулу (85.6) слагаемое того же знака, что и в классическую формулу, во втором — противоположного. В соответствии с этим различают *положительную* и *отрицательную дисперсию*. Ход показателя преломления $n(\omega)$ вблизи собственной частоты $\omega_0 = |\omega_{nk}|$ в обоих случаях представлен схематически на рис. 300 (n_+ для положительной, n_- для отрицательной дисперсии). Что касается коэффициента затухания γ, то в случае отрицательной дисперсии он также *отрицателен*, т.е. при распространении света имеет место его *усиление*, а не ослабление. Это происходит, конечно, за счет переходов атомов с более высокого энергетического уровня на более низкий с испусканием квантов света (индуцированное излучение, см. § 119).

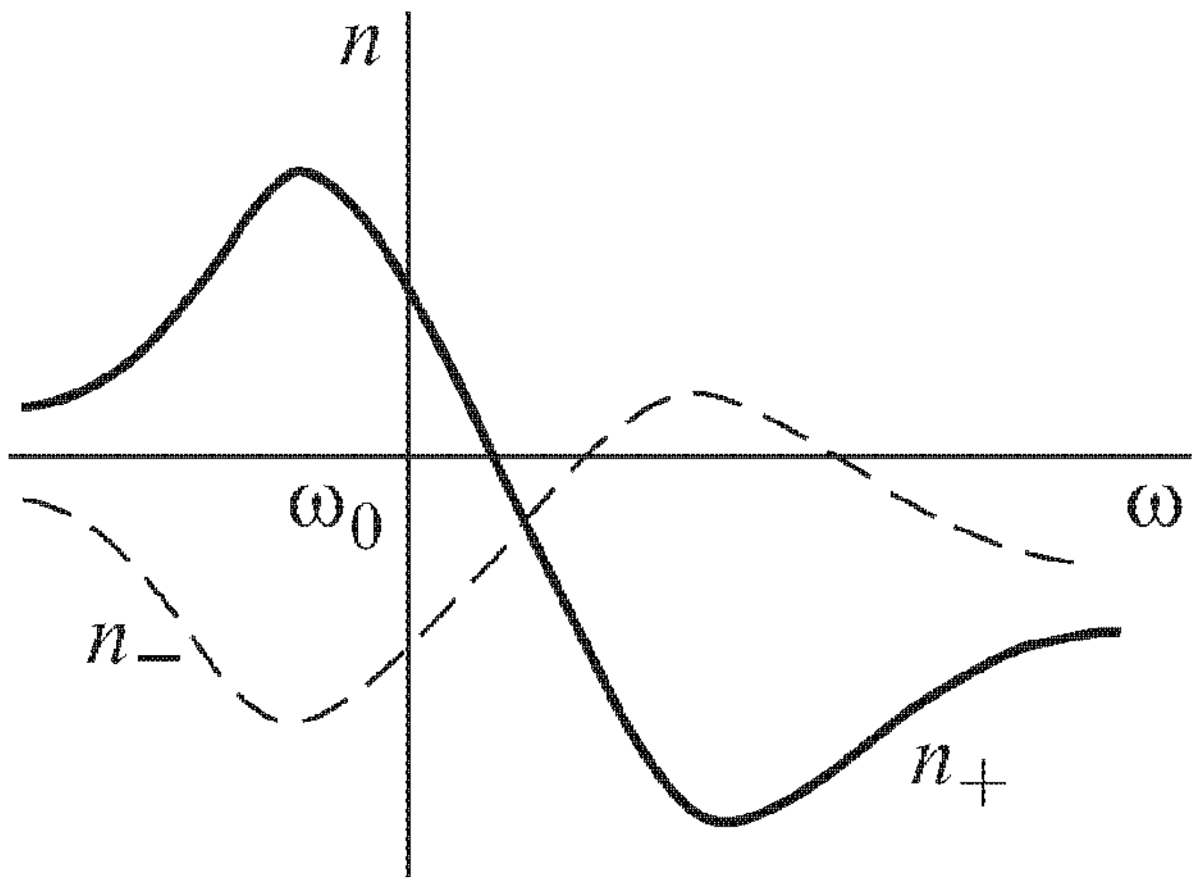


Рис. 300

В обычных условиях до включения внешнего поля атомы среды находятся на самом низком — основном — энергетическом уровне. В этом случае световая волна будет переводить атомы только на более высокие уровни, так что все f_{nk} окажутся положительными. То же самое будет происходить и при тепловом возбуждении в случае термодинамического равновесия, так как, в соответствии с формулой Больцмана, число атомов в исходном состоянии будет убывать с возрастанием номера энергетического уровня. Однако, применяя нетепловые методы возбуждения, например, пропуская через вещество сильные электрические токи, можно создавать *термодинамически неравновесные* метастабильные состояния вещества с *инверсной заселенностью энергетических уровней*, как это в действительности и делается в лазерах. Под инверсной заселенностью двух различных энергетических уровней понимают такое состояние, когда на верхнем уровне находится больше атомов, чем на нижнем. В этом случае можно получать и действительно получают среду с отрицательной дисперсией. Влияние отрицательных слагаемых в дисперсионной формуле впервые (1930 г.) наблюдал Ладенбург в газе при прохождении через него сильного электрического разряда, хотя дисперсия в целом в его опытах и оставалась положительной.

§ 86. Методы экспериментального исследования аномальной дисперсии

1. Аномальная дисперсия впервые наблюдалась Леру (1832–1907) в 1861 г. Он обнаружил, что в призме, наполненной парами йода, синие лучи преломляются меньше красных (промежуточные лучи парами йода сильно поглощаются и ускользают от наблюдения). Отсюда следует, что при переходе от красных к синим лучам показатель преломления паров йода должен убывать, т.е. в указанной области спектра дисперсия паров йода *аномальна.*

Открытие Леру не обратило на себя должного внимания, пока Христиансен (1848–1917) в 1870 г. не обнаружил и детально изучил аномальную дисперсию в растворе фуксина в спирте. Здесь показатель преломления наименьший ($n = 1{,}285$) для фиолетовых и наибольший ($n = 1{,}561$) для желтых лучей, зеленые лучи поглощаются.

Систематические экспериментальные исследования аномальной дисперсии были выполнены Кундтом (1839–1894). Он установил важный экспериментальный результат, что *аномальный ход дисперсии всегда сопровождается поглощением.* После экспериментальных работ Кундта и разработки теории дисперсии стало ясно, что аномальная дисперсия не есть какое-то редкое явление природы, а должна наблюдаться у всех веществ в тех областях спектра, где имеется сильное поглощение.

При изучении аномальной дисперсии Кундт пользовался *методом скрещенных призм*, который применялся еще Ньютоном в его исследованиях по дисперсии света. Идея метода состоит в следующем. Берутся две призмы, из которых первая изготовлена из вещества с *нормальной дисперсией* (обычно из стекла), вторая — из исследуемого вещества. Источником света, как во всяком спектроскопе, служит освещаемая щель, помещаемая в фокальной плоскости коллиматорной линзы. Первая призма, ребро которой устанавливается *вертикально*, разлагает падающий свет в цветную горизонтальную полоску (спектр). Вторая призма, преломляющее ребро которой *горизонтально*, помещается за первой призмой. Она смещает каждую точку полоски вверх или вниз, в зависимости от того, куда обращена эта призма своим основанием: вверх или вниз. Смещение зависит от длины волны. Вследствие этого полоска искривляется и становится наклонной. Если дисперсия второй призмы нормальная, то полоска монотонно поднимается или опускается (рис. 301 *а*). Если же она аномальная, то в результате поглощения лучей с аномальной преломляемостью полоска разрывается на две части, края которых, примыкающие к полосе поглощения, загибаются в противоположные стороны (рис. 301 *б*). Разумеется, для получения описанной картины должна применяться система линз, дающая на экране изображение освещаемой щели в различных цветах спектра.

В изящной лекционной демонстрации Кундта, широко применяющейся до настоящего времени, роль второй призмы играло конусообразное пламя горелки, в которое вводился металлический натрий. В результате прохождения света через пламя горелки в спектре наблюдалась не только темная полоска, соответствующая желтой линии поглощения паров натрия, но и загиб в противоположные стороны разорванных частей спектра, примыкающих к области поглощения (рис. 301 *б*). D-линия поглощения паров натрия в желтой части спектра состоит из двух близко расположенных тонких линий с длинами волн $\lambda_{D_1} = 589{,}0$ и $\lambda_{D_2} = 589{,}6$ нм. В описанной демонстрации плотность паров натрия была настолько велика, что обе линии сливались в одну полоску, и детали явления не разрешались. Их можно разрешить, уменьшая плотность паров натрия и улучшая условия опыта. Тогда можно наблюдать две области аномальной дисперсии, соответствующие линиям поглощения с длинами волн λ_{D_1} и λ_{D_2} (рис. 302).

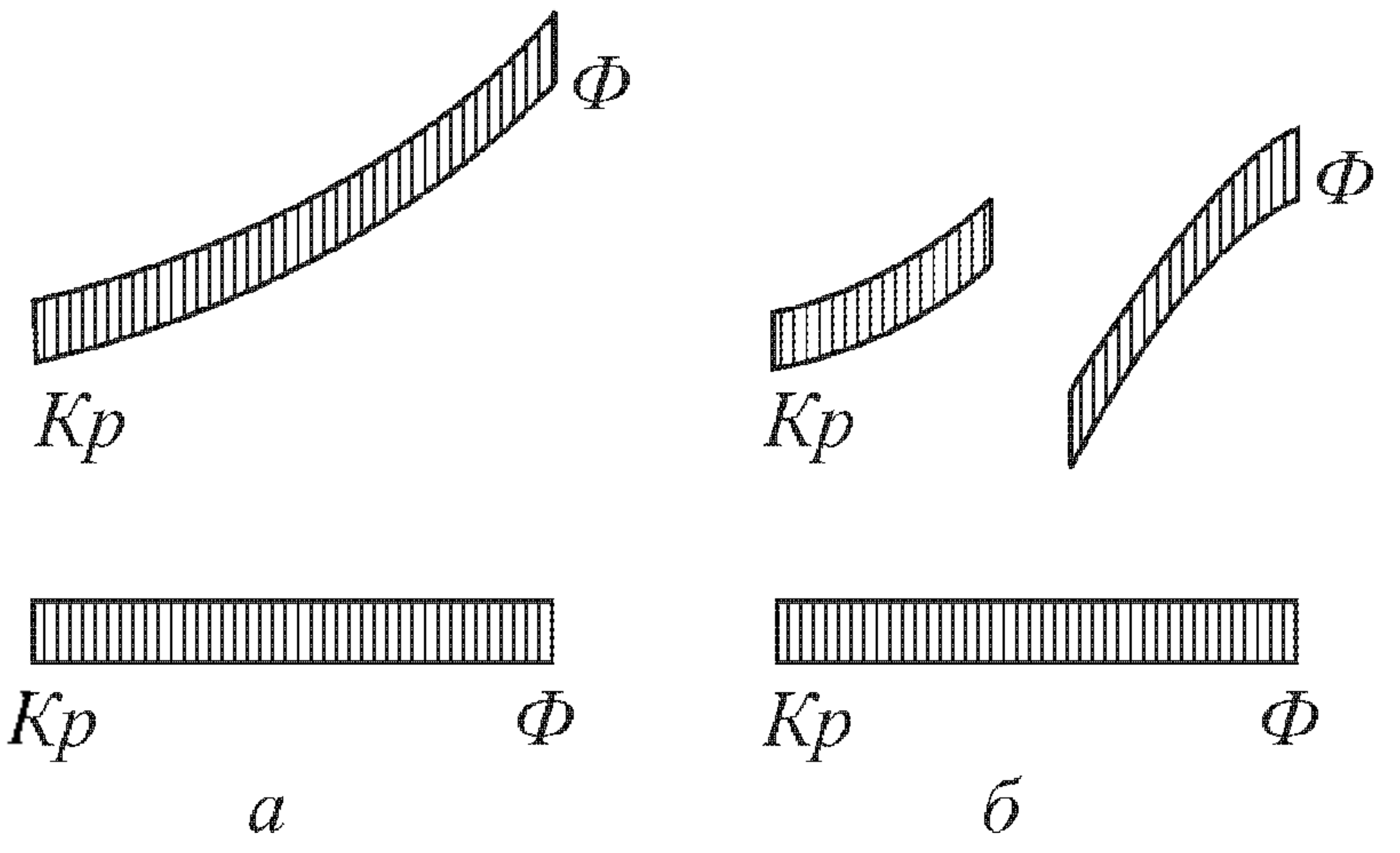


Рис. 301

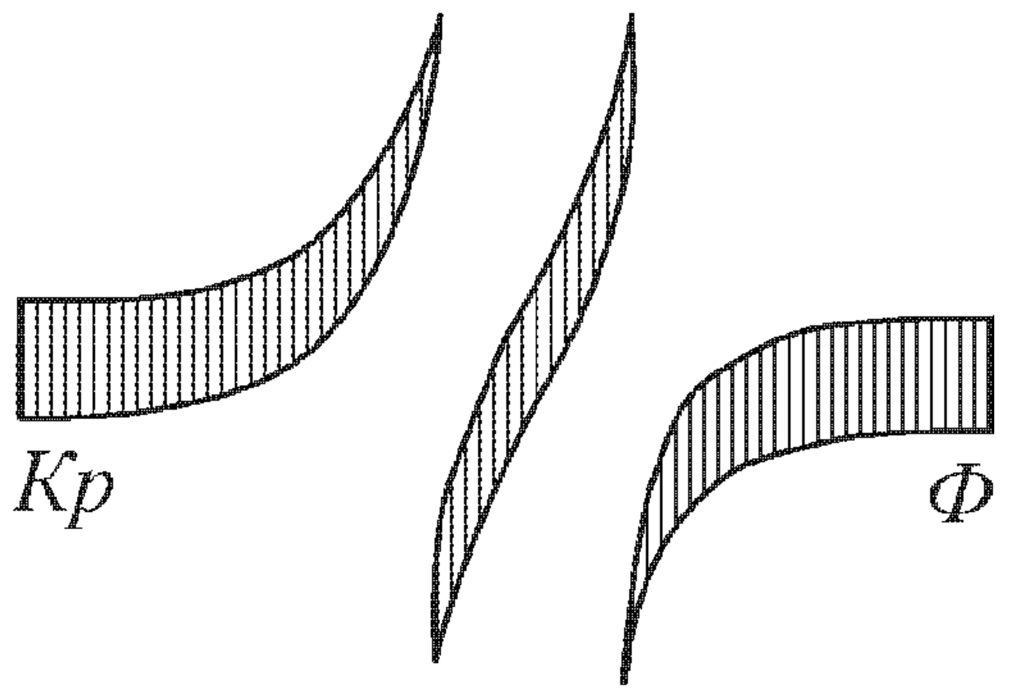


Рис. 302

2. Как уже указывалось в п. 2 § 84, проверку теории лучше всего производить на газах малой плотности, характеризующихся тонкими и резкими линиями поглощения. Однако при малой плотности, необходимой для работы вблизи линии поглощения, показатели преломления газов малы. Для их измерения наилучшим является интерференционный метод, предложенный Пуччианти (1901 г.) и усовершенствованный Д.С. Рождественским в его классических работах по аномальной дисперсии в парах натрия (1912 г.).

В этом методе интерферометр (например, Жамена) скрещивается со спектрографом (дифракционной решеткой или призмой с большой дисперсией). Интерферометр устанавливают так, что он дает в белом свете горизонтальные интерференционные полосы на щели спектрографа, установленной вертикально. Цветную интерференционную картину, полученную на щели, спектрограф развертывает в горизонтальном направлении. Спектр оказывается пересеченным в продольном направлении интерференционными полосами, каждая из которых характеризуется одним и тем же порядком интерференции. Цвет полосы меняется вдоль ее длины от красного к фиолетовому, а расстояния между полосами при этом уменьшаются из-за уменьшения длины волны. Интерференционным минимумам соответствуют темные линии. Таким образом, спектр будет пересечен в продольном направлении темными линиями, сужающимися от красного конца спектра к фиолетовому.

Перед измерением прибор регулируется так, чтобы нулевая интерференционная полоса была прямолинейной и горизонтальной. Примем ее за ось X с направлением от красного конца к фиолетовому, а ось Y направим параллельно щели спектрографа. Ширина интерференционной полосы пропорциональна λ и может быть представлена в виде $\Delta y = \alpha\lambda(x)$, где α — постоянная прибора, практически не зависящая от номера полосы (порядка интерференции) k. Поэтому ордината k-й полосы будет $y_k = k\,\Delta y = \alpha k\lambda(x)$. Это есть уравнение кривой, определяющей форму рассматриваемой полосы.

Введем теперь в одно из плеч интерферометра слой исследуемого вещества толщины l с показателем преломления $n(\lambda)$. Тогда добавится разность хода $(n-1)l$, в результате чего полосы интерференции сместятся вверх или вниз на $(n-1)l/\lambda$ полос, т.е. на расстояние $(n-1)l/\lambda(\Delta y)$. Уравнение, определяющее форму k-й полосы, примет вид

$$y_k = \alpha[k\lambda(x) \pm (n-1)l] \tag{86.1}$$

(знак определяется тем, в какое из плеч интерферометра введен слой исследуемого вещества). В частности, при $k = 0$ получаем $y_0 = \alpha(n-1)l$. Отсюда видно, что нулевая интерференционная полоса вычерчивает в определенном масштабе график зависимости показателя преломления n от абсциссы x, а следовательно, и от длины волны λ, т.е. определяет дисперсию исследуемого вещества. Полосы ненулевого порядка имеют дополнительный наклон, изменяющийся с изменением k (см. рис. 303, заимствованный из работ Рождественского).

3. Небольшое изменение метода позволило Рождественскому значительно повысить точность измерений в окрестности полосы поглощения. Измененный метод получил название *метода крюков*. Допустим, что в одно из плеч интерферометра введено исследуемое вещество (газ или пар), а в другое — стеклянная пластинка толщины $l_{\text{ст}}$ с показателем преломления $n_{\text{ст}}$. Пластинка вносит между интерферирующими пучками разность хода $(n_{\text{ст}} - 1)l_{\text{ст}}$, смещая интерференционную картину вверх или вниз на расстояние $(n_{\text{ст}} - 1)l_{\text{ст}}\alpha$. Теперь ордината k-й полосы будет определяться выражением

$$y_k = \alpha[k\lambda(x) \pm (n-1)l \mp (n_{\text{ст}} - 1)l_{\text{ст}}]. \tag{86.2}$$

Интерференционная полоса нулевого порядка уйдет из поля зрения. Ее место в поле зрения займет другая интерференционная полоса, порядок k которой найдется, если в (86.2) положить $y_k = 0$. Это дает

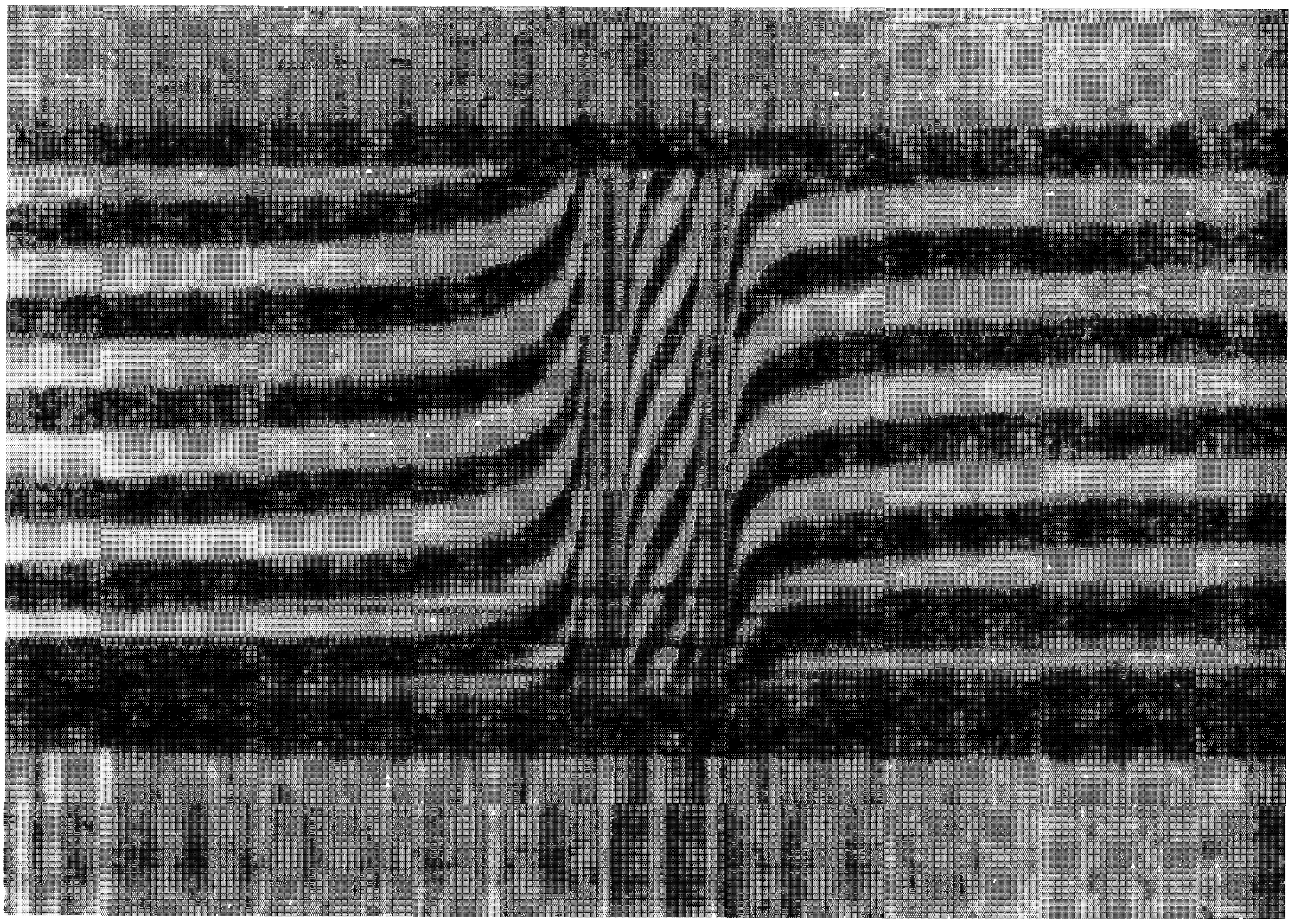


Рис. 303

$$k = \pm\frac{n_{\text{ст}} - 1}{\lambda}\, l_{\text{ст}} \mp \frac{n-1}{\lambda}\, l. \tag{86.3}$$

Из-за близости к единице показателей преломления газов последнее слагаемое мало и при вычислении k может не учитываться. Таким путем для пластинки с $l_{\text{ст}} = 1$ мм, $n_{\text{ст}} = 1{,}5$ при длине волны $\lambda = 600$ нм получаем $k \sim 10^4$. В белом свете интерференция столь высокого порядка невозможна. Интерференция в белом свете становится возможной благодаря наличию спектроскопа, играющего роль монохроматора, пространственно разлагающего белый свет на квазимонохроматические составляющие. Поэтому длина волны λ вдоль каждой интерференционной полосы изменяется, т.е. является функцией координаты x.

Дифференцируя y_k по x при постоянном k, находим наклон k-й интерференционной полосы:

$$\frac{dy_k}{dx} = \alpha\frac{d\lambda}{dx}\left(k \pm l\frac{dn}{d\lambda} \mp l_{\text{ст}}\frac{dn_{\text{ст}}}{d\lambda}\right). \tag{86.4}$$

Благодаря малой дисперсии стекла последнее слагаемое мало и может быть отброшено. Если в отсутствие стеклянной пластинки и исследуемого вещества нулевая интерференционная полоса горизонтальна, то внесение одной только пластинки делает интерференционные полосы наклонными. Наклон полосы определяется производной $\frac{dy_k}{dx} = k\alpha\frac{d\lambda}{dx}$, т.е. он тем больше, чем больше порядок интерференции k. Наклон, вызываемый стеклянной пластинкой, происходит в одну и ту же сторону во всей исследуемой области спектра. При перенесении пластинки в другое плечо интерферометра знак k, а с ним и наклон интерференционных полос меняются на противоположные. Поместим пластинку в то плечо интерферометра, чтобы было $k > 0$, т.е.

$$k \approx \frac{n_{\text{ст}} - 1}{\lambda}\, l_{\text{ст}}. \tag{86.5}$$

Тогда при нашем выборе положительного направления оси X ($d\lambda/dx < 0$) получится $dy_k/dx < 0$, т.е. полосы интерференции будут наклонены сверху (от положительного конца оси Y) вниз направо (в сторону положительного — фиолетового — конца оси X).

Допустим теперь, что во второе плечо интерферометра введен исследуемый газ. Показатель преломления газа вдали от линии поглощения очень близок к единице и на наклон полос практически не оказывает никакого влияния. Зато вблизи линии поглощения велика производная $dn/d\lambda$, вызывающая сильный загиб интерференционных полос. В отсутствие пластинки наклон будет определяться производной

$$\frac{dy_k}{dx} = -\alpha l\frac{d\lambda}{dx}\,\frac{dn}{d\lambda}.$$

Следовательно, в области нормальной дисперсии ($dn/d\lambda < 0$) получится $dy_k/dx > 0$, т.е. в этом случае интерференционные полосы пойдут сверху вниз налево.

Таким образом, в области нормальной дисперсии стеклянная пластинка и исследуемый газ вызывают наклон полос в противоположные стороны. При одновременном действии газа и стеклянной пластинки с надлежаще подобранной толщиной эти действия в определенной точке интерференционной полосы могут компенсировать друг друга. В такой точке касательная к интерференционной полосе станет горизонтальной, т.е. на полосе получится *крюк*, по выражению Рождественского. Вблизи линии поглощения таких крюков получится два. Оба они расположены в области нормальной дисперсии. Слева (со стороны длинных волн) крюк будет обращен вершиною вниз, а справа (со стороны коротких волн) — вверх (рис. 304). В вершине крюка должно быть $dy_k/dx = 0$, или $dy_k/d\lambda = 0$, т.е.

$$k + l\frac{dn}{d\lambda} - l_{\text{ст}}\frac{dn_{\text{ст}}}{d\lambda} = 0,$$

или

$$l\frac{dn}{d\lambda} = -k + l_{\text{ст}}\frac{dn_{\text{ст}}}{d\lambda}. \qquad (86.6)$$

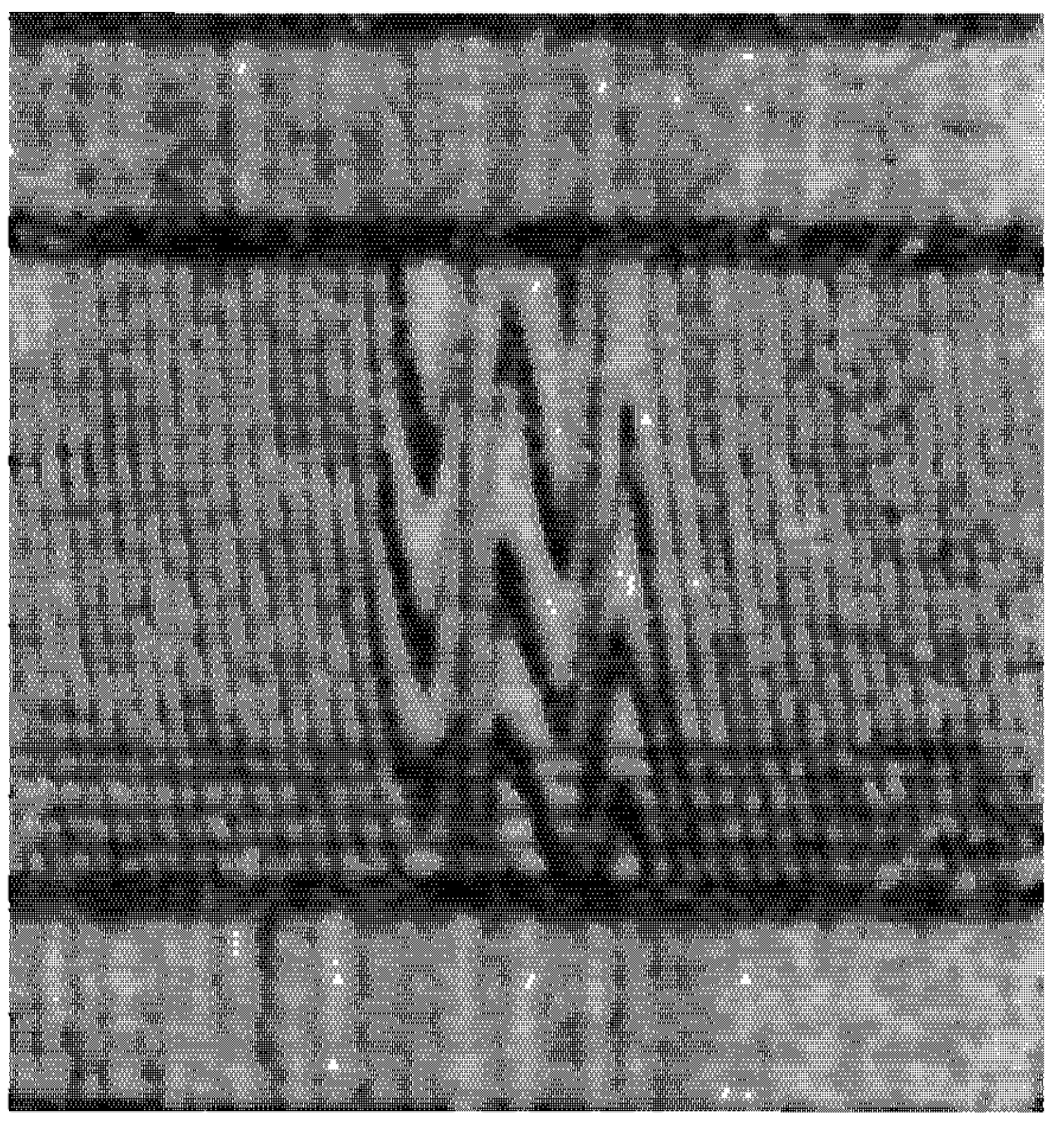


Рис. 304

Последнее слагаемое, как уже указывалось, пренебрежимо мало, а первое может быть вычислено по формуле (86.5). Таким образом, по формуле (86.6) можно вычислить значения дисперсии газа $dn/d\lambda$ для тех значений λ, которые соответствуют вершинам крюков, т.е. точкам загиба интерференционных полос.

Уже в начале своих исследований Рождественский убедился, что вдали от линии поглощения формула Зельмейера правильно передает ход показателя преломления в зависимости от длины волны. Эту формулу следует писать в виде

$$n^2 = 1 + 4\pi\frac{Nfe^2/m}{\omega_0^2 - \omega^2} = 1 + \frac{Nf\lambda_0^2\lambda^2e^2/m}{\pi c^2(\lambda^2 - \lambda_0^2)}, \qquad (86.7)$$

где f — сила осциллятора. Обозначим через $\Delta\lambda$ расстояние вершины одного из крюков от линии поглощения λ_0 (тогда расстояние между вершинами обоих крюков будет $2\Delta\lambda$). Вычислим по формуле (86.7) производную $dn/d\lambda$, учитывая при этом, что $|\Delta\lambda| \ll \lambda_0$ и $n - 1 \ll 1$. Тогда из формулы (86.6), пренебрегая последним слагаемым, найдем

$$f = \frac{4\pi c^2 k}{Nl\lambda_0^3e^2/m} = \frac{4\pi c^2(n_{\text{ст}} - 1)l_{\text{ст}}}{Nl\lambda_0^4le^2/m}(\Delta\lambda)^2. \qquad (86.8)$$

Таким образом, силу осциллятора можно найти по расстоянию между вершинами крюков.

§ 87. Дисперсия плазмы

1. Плазма есть ионизованный газ, в котором электроны и ионы могут рассматриваться как свободные частицы с собственными частотами, равными нулю (см. т. III, § 121). Диэлектрическая проницаемость плазмы определяется в основном *свободными электронами.* Влиянием ионов можно пренебречь, так как их массы практически бесконечно велики по сравнению с массами легких электронов. Полагая в формуле (84.5) $\omega_0 = 0$ и пренебрегая затуханием, получим для плазмы

$$\varepsilon = 1 - \left(\frac{\omega_p}{\omega}\right)^2, \tag{87.1}$$

где введено обозначение

$$\omega_p^2 = 4\pi N e^2/m, \tag{87.2}$$

а N означает концентрацию свободных электронов. Величина ω_p называется *плазменной* или *ленгмюровской частотой.* Частота ω_p играет для плазмы роль *собственной частоты.* Однако она характеризует не отдельные частицы, а *весь коллектив заряженных частиц,* из которых состоит плазма.

Чтобы понять, как может появиться такая «собственная частота» у коллектива частиц, каждая из которых в отдельности собственными частотами не обладает, рассмотрим следующий пример. Допустим, что нейтральная плазма занимает пространство между бесконечными плоскостями, перпендикулярными к оси X. Среднее электрическое поле в такой нейтральной плазме равно нулю. Сместим все электроны плазмы параллельно оси X на малое расстояние x (рис. 305), а ионы оставим несмещенными. Тогда на границах плазмы возникнут электрические заряды с поверхностной плотностью $\sigma = Nex$, и в плазме возникнет электрическое поле $E = 4\pi\sigma = 4\pi Nex$. На каждый электрон будет действовать квазиупругая сила $F = 4\pi Ne^2 x$. Если плазму предоставить самой себе, то возникнет свободное гармоническое колебание электронов с собственной частотой $\sqrt{4\pi Ne^2/m}$. Но это и есть плазменная частота.

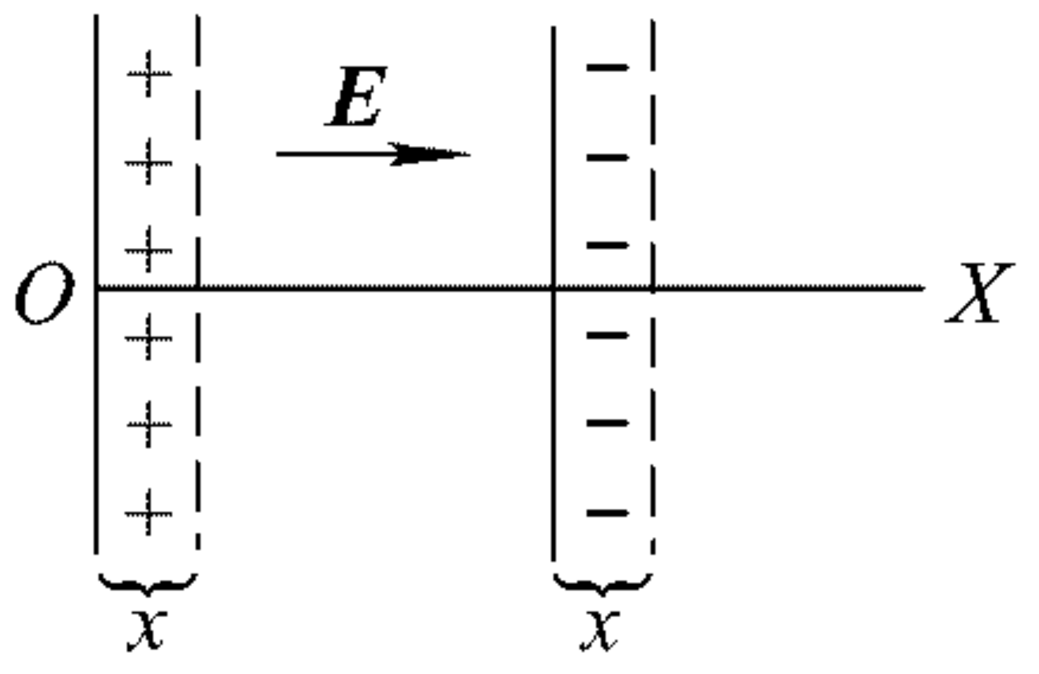


Рис. 305

2. Для плазменной частоты $\omega = \omega_p$ диэлектрическая проницаемость ε обращается в нуль. При $\omega > \omega_p$ величина ε (а с ней и показатель преломления $n = \sqrt{\varepsilon}$) положительна, но меньше единицы. При $\omega < \omega_p$ диэлектрическая проницаемость ε отрицательна, а показатель преломления чисто мнимый, т.е. $n = -i\varkappa$. Поэтому длинные электромагнитные волны (частота которых $\omega < \omega_p$) в плазме распространяться не могут. Они могут проникать только в тонкий поверхностный слой плазмы, испытывая от него полное отражение. Действительно, предположим, что падающая волна поляризована перпендикулярно к плоскости падения. (Случай другой поляризации разбирается так же.)

Тогда по формуле Френеля

$$\frac{R}{\mathcal{E}} = \frac{\cos\varphi - \sqrt{\varepsilon}\cos\psi}{\cos\varphi + \sqrt{\varepsilon}\cos\psi} = \frac{\cos\varphi + i\varkappa\cos\psi}{\cos\varphi - i\varkappa\cos\psi},$$

причем $\sin\varphi/\sin\psi = \sqrt{\varepsilon} = -i\varkappa$. Следовательно, $\cos\psi = \sqrt{1-\sin^2\psi} = \sqrt{1+\sin^2\varphi/\varkappa^2}$. Таким образом, $\cos\psi$ — величина вещественная, а потому $|R/\mathcal{E}| = 1$, что и доказывает наше утверждение.

Изложенное играет исключительно важную роль в осуществлении на Земле *дальней радиосвязи.* В земной атмосфере имеется ионизованная область, называемая *ионосферой.* Она начинается примерно с высоты 60 км и простирается, по-видимому, до высот $\sim$ 20 000 км. Основными источниками ионизации ионосферы являются ультрафиолетовое излучение Солнца и мягкое (от 0,8 до 30 нм) рентгеновское излучение солнечной короны. Другим источником служит корпускулярное излучение Солнца. Концентрация электронов N меняется с высотой неравномерно. Имеется несколько относительных максимумов ионизации, расположенных на различных высотах. Область ионосферы, содержащая один из таких максимумов, условно называется *ионосферным слоем.* Слои, расположенные в порядке возрастания высоты, обозначаются через D, E_1, E_2, F_1, F_2. Максимумы электронной концентрации в них меняются примерно в пределах $10^4 - 10^6$ электронов на 1 см^3. Концентрация электронов зависит от географической широты места и испытывает регулярные суточные и годичные изменения. Летом она больше, чем зимой, днем больше, чем ночью. Кроме того, наблюдаются *спорадические* изменения концентрации, вызванные вспышками на Солнце и пр.

Посмотрим теперь, как радиоволны, излученные какой-либо радиостанцией A, находящейся на земной поверхности, могут достигнуть приемника B, расположенного также на земной поверхности на расстоянии нескольких тысяч километров. Прямой путь через землю исключен, так как радиоволны в земле сильно поглощаются из-за ее высокой электрической проводимости. Показатель преломления неионизованного воздуха очень мало отличается от единицы, так что рефракция радиоволн практически не играет роли. Если бы не было ионосферы, то единственным способом достигнуть приемника B была бы дифракция. Но приемник B расположен глубоко в области геометрической тени на расстоянии в тысячи или десятки тысяч длин волн от ее границы. При таких условиях интенсивность дифрагированной волны в точке нахождения приемника B будет ничтожно мала, и никакой приемник практически не сможет обнаружить эту волну. Положение меняется при наличии ионосферы, так как радиоволна может *отразиться от ионосферы* и таким путем достигнуть приемника. Только благодаря такому отражению возможна передача радиосигналов на земной поверхности на многие тысячи километров.

3. Найдем связь между фазовой v и групповой u скоростями электромагнитных волн в плазме при $\omega_p > \omega$. Используя выражение (87.1), для волнового числа k получаем

$$c^2k^2 = \omega^2\varepsilon = \omega^2 - \omega_p^2.$$

Дифференцирование этого соотношения дает: $c^2k\,dk = \omega\,d\omega$, т.е. $(\omega/k)(d\omega/dk) = c^2$, или

$$vu = c^2. \tag{87.3}$$

Фазовая скорость в плазме

$$v = \frac{c}{\sqrt{\varepsilon}} = \frac{c}{\sqrt{1-\omega_p^2/\omega^2}} \tag{87.4}$$

всегда больше скорости света в вакууме. Для групповой скорости соотношение (87.3) дает

$$u = c^2/v = c\sqrt{1-\omega_p^2/\omega^2}. \tag{87.5}$$

Она всегда меньше c, как это и должно быть.

Отметим интересное астрофизическое применение формулы (87.5). После открытия Хьюишем в 1967 г. *пульсаров* (*нейтронных звезд*) сразу же было обнаружено, что *длинноволновые сигналы доходят от пульсаров до Земли медленнее коротковолновых.* (В этом можно убедиться, принимая один и тот же сигнал с помощью двух радиоприемников, настроенных на разные частоты.) Это было объяснено влиянием *межзвездной плазмы*, через которую проходит сигнал. Квазимонохроматический сигнал распространяется в межзвездной плазме с групповой скоростью (87.5). Время распространения сигнала от пульсара до Земли определяется интегралом $t = \int dx/u$ по всему пути сигнала. Концентрация свободных электронов N, а с ней и плазменная частота ω_p имеют разные значения в разных точках пути. Однако всюду $\omega_p \ll \omega$, так что можно ограничиться первым членом разложения подынтегрального выражения по степеням отношения ω_p^2/ω^2. Это дает

$$t = \frac{1}{c}\int\left(1-\omega_p^2/\omega^2\right)^{-1/2}dx = \frac{1}{c}\int\left(1+\frac{1}{2}\frac{\omega_p^2}{\omega^2}\right)dx.$$

По сравнению с вакуумом время распространения сигнала увеличивается на

$$\Delta t = \frac{1}{2c\omega^2}\int\omega_p^2\,dx = \frac{\lambda^2e^2}{2\pi cmc^2}\int N\,dx = \frac{\lambda^2 r}{2\pi c}\int N\,dx,$$

где $r = e^2/(mc^2) \approx 2{,}8 \cdot 10^{-13}$ см — *классический радиус электрона.*

1. Интеграл $\int N\,dx$ имеет смысл полного числа электронов в цилиндрическом канале, поперечное сечение которого равно 1 см^2, а длина — пути, пройденному сигналом от пульсара до Земли. Он является одной из интегральных характеристик межзвездной плазмы на пути распространения сигнала. Несмотря на ничтожную концентрацию такой плазмы, из-за колоссальности расстояний до пульсаров величина этого интеграла оказалась достаточной, чтобы обнаружить запаздывание (в области сантиметровых волн) длинноволновых сигналов относительно коротковолновых. Таким путем впервые были оценены расстояния до пульсаров. В предположении, что на пути от пульсаров к Земле около 10 % атомов водорода ионизовано, было найдено, что расстояния до большинства зарегистрированных пульсаров лежат между 200 и 7000 световых лет.

§ 88. Средняя плотность электромагнитной энергии в диспергирующих средах

1. Выражение для плотности электромагнитной энергии $w = (\varepsilon E^2 + \mu H^2)/(8\pi)$ получается в предположении, что ε и μ *постоянны*, т.е. не зависят от частоты ω (см. т. III, § 84). *В случае диспергирующих сред это выражение неприменимо.* Не разбирая этот вопрос в общем виде, выведем выражение для средней плотности электромагнитной энергии в непоглощающей диспергирующей среде на частном примере, принадлежащем М.Л. Левину.

Пусть вещество с диэлектрической проницаемостью $\varepsilon(\omega)$ и магнитной проницаемостью $\mu(\omega)$ заполняет плоский конденсатор с емкостью $C = \varepsilon(\omega)C_0$ и тонкий соленоид с индуктивностью $L = \mu(\omega)L_0$, соединенные в колебательный контур (рис. 306). Здесь C_0 и L_0 — значения емкости и индуктивности для того случая, когда в пространстве между обкладками конденсатора и внутри соленоида вакуум. При отсутствии сопротивления в контуре будут совершаться свободные гармонические колебания с циклической частотой $\omega = 1/\sqrt{L(\omega)C(\omega)}$. Если в некоторый момент времени ввести в контур малое сопротивление R, то, начиная с этого момента, колебания сделаются затухающими и первоначально запасенная электромагнитная энергия будет переходить в джоулево тепло, выделяющееся в сопротивлении R. Полное количество тепла, выделившееся в сопротивлении R за время, когда колебания прекратятся, будет равно электромагнитной энергии, запасенной в контуре до введения сопротивления. Поэтому задача сводится к вычислению джоулева тепла.

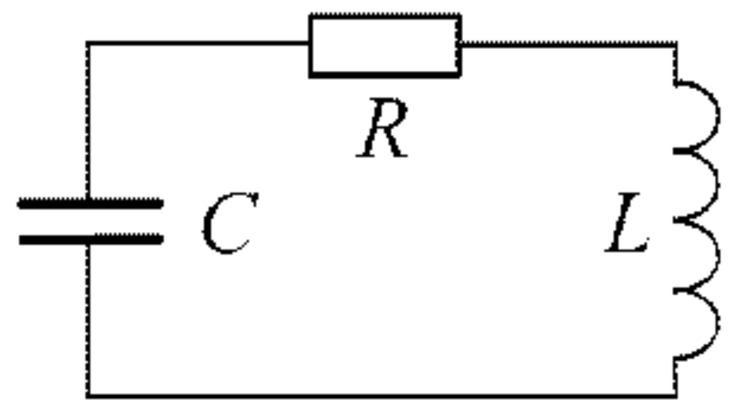


Рис. 306

Пусть при $t < 0$ в контуре совершаются свободные колебания:

$$I = I_0 e^{i\omega t}, \quad V = V_0 e^{i\omega t}.$$

где I — сила тока в контуре, а V — напряжение на обкладках конденсатора, связанные между собой соотношением $L\dot{I}+V=0$, или $i\omega LI+V=0$. Если в момент $t=0$ в контур ввести сопротивление R, то, начиная с этого момента, колебания будут описываться уравнением

$$L(\widetilde{\omega})\ddot{I}+R\dot{I}+\frac{I}{C(\widetilde{\omega})}=0,$$

откуда

$$I=I_0e^{i\widetilde{\omega}t},\quad t>0,$$

где $\widetilde{\omega}$ — комплексная частота, определяемая уравнением

$$\widetilde{\omega}L(\widetilde{\omega})-\frac{1}{\widetilde{\omega}C(\widetilde{\omega})}=iR.$$

Если R исчезающе мало, то $\widetilde{\omega}$ должна отличаться от ω также на исчезающе малую величину. Но ω удовлетворяет уравнению

$$\omega L(\omega)-\frac{1}{\omega C(\omega)}=0.$$

Вычитая его из предыдущего соотношения и заменяя все разности дифференциалами, получим

$$\left[\frac{d}{d\omega}(\omega L)+\frac{1}{\omega^2C^2}\frac{d}{d\omega}(\omega C)\right](\widetilde{\omega}-\omega)=iR,$$

откуда $\widetilde{\omega}=\omega+i\delta$, причем

$$\frac{R}{\delta}=\frac{d(\omega L)}{d\omega}+\frac{1}{\omega^2C^2}\frac{d(\omega C)}{d\omega}=\frac{d(\omega L)}{d\omega}+\frac{L}{C}\frac{d(\omega C)}{d\omega}.$$

Для определения джоулева тепла надо проинтегрировать выражение RI^2 по времени. Поскольку возведение в квадрат — нелинейная операция, необходимо перейти к вещественной форме, т.е. сделать замену

$$I\to\mathrm{Re}\,(I)=(I+I^*)/2.$$

Энергия, первоначально запасенная в колебательном контуре, равна

$$W=\int\limits_0^\infty R\left(\frac{I+I^*}{2}\right)^2dt=\frac{R|I_0|^2}{4}\left\{\frac{\delta}{\omega^2+\delta^2}+\frac{1}{\delta}\right\}$$

или в пределе при $\delta\to 0$

$$W=\frac{|I_0|^2}{4}\frac{R}{\delta}.$$

Подставляя сюда значение для R/δ и пользуясь соотношением $\omega L|I_0|=|V_0|$, получим

$$W=\frac{L_0|I_0|^2}{4}\frac{d(\omega\mu)}{d\omega}+\frac{C_0|V_0|^2}{4}\frac{d(\omega\varepsilon)}{d\omega}.$$

Если бы между обкладками конденсатора и внутри соленоида был вакуум, то для средних по времени значений магнитной и электрической энергий можно было бы написать

$$\frac{L_0|I_0|^2}{4}=\frac{1}{8\pi}\overline{\mathbf{H}^2}\tau_m,\qquad\frac{C_0|V_0|^2}{4}=\frac{1}{8\pi}\overline{\mathbf{E}^2}\tau_e,$$

где τ_m и τ_e — объемы соленоида и конденсатора, а **E** и **H** — напряженности электрического и магнитного полей, когда амплитуды напряжения на конденсаторе и тока в соленоиде равны V_0 и I_0. Но при заданных V_0 и I_0 поля **E** и **H** не зависят от среды, заполняющей конденсатор и соленоид. Поэтому предыдущие соотношения остаются справедливыми и в том случае, когда конденсатор и соленоид заполнены веществом. Используя их, получаем следующие выражения для средних по времени значений плотностей электрической и магнитной энергий:

$$\begin{aligned} \overline{w}_e &= \frac{\overline{W}_e}{\tau_e} = \frac{1}{8\pi}\frac{d(\omega\varepsilon)}{d\omega}\overline{\mathbf{E}^2}, \\ \overline{w}_m &= \frac{\overline{W}_m}{\tau_m} = \frac{1}{8\pi}\frac{d(\omega\mu)}{d\omega}\overline{\mathbf{H}^2}. \end{aligned} \tag{88.1}$$

Принципиальный недостаток приведенного вывода состоит в том, что в нем дифференцирование функций ωL и ωC производится вдоль *мнимой оси* (так как разность частот $\widetilde{\omega} - \omega = iR$ — величина чисто мнимая), а в окончательном выражении (88.1) производится подмена дифференцированием по *вещественной переменной* ω. Так можно поступать, когда функции ωL и ωC *аналитичны*. Поэтому для полноты доказательства надо было бы доказать аналитичность этих функций, чего в выводе Левина нет. Это можно сделать в общей теории дисперсии, исследуя аналитические свойства функций $\varepsilon(\omega)$ и $\mu(\omega)$. Однако рассмотрение этого вопроса выходит за рамки нашей книги.

2. Смысл формулы (88.1) полезно уяснить на примере газа классических гармонических осцилляторов в монохроматическом электрическом поле с частотой ω. Вдали от собственной частоты ω_0 осциллятора можно пренебречь затуханием. Тогда смещение осциллятора из положения равновесия выразится формулой

$$\mathbf{r} = \frac{e\mathbf{E}}{m(\omega_0^2 - \omega^2)}.$$

В этом случае энергия слагается из *энергии самого электромагнитного поля* (т.е. поля в вакууме) и из *энергии частиц*, находящихся в поле. Последняя энергия в свою очередь состоит из *кинетической* и *потенциальной энергии* колеблющихся осцилляторов. В статических полях кинетической энергии нет. Это приводит к формуле $w = \varepsilon E^2/(8\pi)$. В переменных полях кинетическую энергию надо учитывать, что и делается ниже.

Плотность собственно энергии электрического поля равна

$$w_1 = \frac{1}{8\pi}\left(\frac{\mathbf{E}+\mathbf{E}^*}{2}\right)^2 = \frac{\mathbf{E}^2}{32\pi} + \frac{\mathbf{E}\mathbf{E}^*}{32\pi} + \text{компл. сопр.}$$

Плотность потенциальной энергии:

$$w_2 = \frac{Nm\omega_0^2}{2}\left(\frac{\mathbf{r}+\mathbf{r}^*}{2}\right)^2 = \frac{Nm\omega_0^2}{8}(\mathbf{r}^2 + \mathbf{r}\mathbf{r}^*) + \text{компл. сопр.}$$

Плотность кинетической энергии:

$$w_3 = \frac{Nm}{2}\left(\frac{\dot{\mathbf{r}}+\dot{\mathbf{r}}^*}{2}\right)^2 = -\frac{Nm\omega^2}{8}(\mathbf{r}^2 - \mathbf{r}\mathbf{r}^*) + \text{компл. сопр.}$$

Подставляя сюда выражение для **r** и замечая, что из формулы Зельмейера следует

$$\frac{d(\omega\varepsilon)}{d\omega} = 1 + \frac{(\varepsilon-1)(\omega_0^2+\omega^2)}{\omega_0^2-\omega^2},$$

получим для плотности электрической энергии:

$$w_e = \frac{\varepsilon \mathbf{E}^2}{32\pi} + \frac{1}{32\pi}\frac{d(\omega\varepsilon)}{d\omega}(\mathbf{E}\mathbf{E}^*) + \text{компл. сопр.}$$

Усредняя по времени, получаем первую формулу (88.1). Для плотности магнитной энергии имеем обычное выражение, как в недиспергирующей среде.

ЗАДАЧИ

1. Рэлей предложил определять *среднюю скорость движения энергии u* в плоской бегущей волне как отношение средней плотности потока энергии к средней плотности самой энергии. Пользуясь выражением для вектора Пойнтинга, показать, что так определенная скорость в случае монохроматической электромагнитной волны совпадает с групповой скоростью.

Решение. Для средних плотностей энергии и ее потока нетрудно получить

$$\overline{w} = \frac{c}{8\pi}\sqrt{\frac{\varepsilon}{\mu}}\frac{dk}{d\omega}(\mathbf{E}\mathbf{E}^*), \qquad \overline{S} = \frac{c}{8\pi}\sqrt{\frac{\varepsilon}{\mu}}(\mathbf{E}\mathbf{E}^*),$$

откуда и следует требуемый результат.

2. Показать, что если $\varepsilon(\omega)$ и $\mu(\omega)$ положительны, то фазовая и групповая скорости в электромагнитной волне направлены в одну сторону.

Решение. Средняя плотность электромагнитной энергии

$$\overline{w} = \frac{1}{8\pi}\frac{d(\omega\varepsilon)}{d\omega}(\mathbf{E}\mathbf{E}^*) + \frac{1}{8\pi}\frac{d(\omega\mu)}{8\pi}(\mathbf{H}\mathbf{H}^*)$$

— существенно положительная величина. В плоской волне $\varepsilon(\mathbf{E}\mathbf{E}^*) = \mu(\mathbf{H}\mathbf{H}^*)$. Поэтому

$$\frac{d(\omega\varepsilon)}{d\omega} + \frac{\mu}{\varepsilon}\frac{d(\omega\mu)}{d\omega} > 0.$$

Это неравенство должно соблюдаться для любых сред, у которых знаки ε и μ совпадают, поскольку оно выведено в предположении, что в среде может распространяться однородная монохроматическая волна, для которой $k^2 = \varepsilon\mu\omega^2/c^2 > 0$. В этом же предположении имеет смысл говорить о групповой скорости. Преобразовав предыдущее неравенство к виду

$$\mu\frac{\omega}{k}\frac{d\omega}{dk} = \mu v u > 0,$$

легко получить требуемый результат.

§ 89. Поглощение света и уширение спектральных линий

1. В классической теории дисперсии поглощение (затухание) излучения учитывается формально с помощью тормозящей силы $-g\dot{\mathbf{r}} = -2m\gamma\mathbf{v}$ в уравнении (84.1). Благодаря этому амплитуда колебаний убывает во времени экспоненциально по закону $\exp(-\gamma t)$, а энергия колебаний — по закону

$$\mathcal{E} = \mathcal{E}_0 e^{-2\gamma t} = \mathcal{E}_0 e^{-t/\tau}. \qquad (89.1)$$

Физическая природа тормозящей силы $-g\mathbf{v}$ при этом остается нераскрытой.

Планк развил теорию дисперсии и поглощения света в предположении, что колеблющийся осциллятор (электрон), двигаясь ускоренно, непрерывно теряет энергию на излучение. Эта убыль энергии на излучение определяется формулой

$$-\frac{d\mathcal{E}}{dt} = \frac{2e^2}{3c^3}\dot{\mathbf{v}}^2 \tag{89.2}$$

(см. т. III, § 141). Применим эту формулу к осциллятору. Если потеря энергии осциллятора за период колебаний относительно мала, то колебания будут отличаться от гармонических мало. Тогда можно считать, что средние за период колебаний значения кинетической и потенциальной энергий осциллятора одинаковы, а потому среднее значение его полной энергии равно удвоенному среднему значению кинетической энергии. Но полная энергия $\mathcal{E}$ слабо затухающего осциллятора в течение периода колебаний остается почти постоянной, так что $\overline{\mathcal{E}} \approx \mathcal{E}$. В случае слабо затухающих колебаний $\mathbf{v} \approx \mathbf{v}_0 \sin(\omega_0 t + \delta)$, $\dot{\mathbf{v}} = \omega_0 \mathbf{v}_0 \cos(\omega t + \delta)$. Возведя последнее соотношение в квадрат и усреднив по периоду колебаний, получим

$$\langle \dot{\mathbf{v}}^2 \rangle = \omega_0^2 \langle \mathbf{v}^2 \rangle = \omega_0^2 \mathcal{E}/m.$$

Усреднив теперь по времени формулу (89.2), найдем

$$-\frac{d\mathcal{E}}{dt} = \frac{2e^2\omega_0^2}{3mc^3}\mathcal{E}.$$

Отсюда снова получается формула вида (89.1). Таким образом, излучение проявляется так, как если бы на колеблющийся заряд действовала тормозящая сила $-2m\gamma\mathbf{v}$, причем

$$\gamma = \frac{1}{3}\frac{e^2\omega_0^2}{mc^3}. \tag{89.3}$$

Затухание, обусловленное излучением, называется *естественным затуханием*. Время естественного затухания $\tau_{\text{изл}}$, в течение которого энергия убывает в e раз, равно

$$\tau_{\text{изл}} = \frac{1}{2\gamma} = \frac{3mc^3}{2e^2\omega_0^2} = \frac{3mc}{8\pi^2 e^2}\lambda^2 = \frac{3}{8\pi^2}\left(\frac{\lambda}{r}\right)\left(\frac{\lambda}{c}\right), \tag{89.4}$$

где $r = e^2/(mc^2) = 2{,}8 \cdot 10^{-13}$ см — *классический радиус электрона*. Для длины волны $\lambda = 600$ нм эта формула дает $\tau_{\text{изл}} = 1{,}63 \cdot 10^{-8}$ с.

2. Затухание делает колебания немонохроматическими. Это приводит к *уширению спектральных линий*. Ширина спектральной линии, обусловленная затуханием вследствие излучения, называется *естественной шириной спектральной линии*. Для ее вычисления поле излучения $E = E_0 \exp(-\gamma t)\sin\omega_0 t$ надо разложить в интеграл Фурье. Квадраты коэффициентов этого разложения определяют *относительные спектральные плотности излучения* $I(\omega)$, т.е. распределение энергии в спектральной линии по спектру частот. Результат выражается ранее выведенной формулой (29.10), которую здесь мы запишем в виде

$$I(\omega) = \frac{1}{(\omega-\omega_0)^2 + 1/(2\tau_{\text{изл}})^2}, \tag{89.5}$$

заменив прежнее время τ на $2\tau_{\text{изл}}$. (Прежнее время τ определяло *затухание по амплитуде*, а здесь через $\tau_{\text{изл}}$ обозначено *время затухания по энергии.*) Форма спектральной линии приведена на рис. 307 для значения параметра $\omega_0\tau_{\text{изл}} = 5$, которое, конечно, слишком мало. (В приведенном выше численном примере $\omega_0\tau_{\text{изл}} = 2\pi c\tau_{\text{изл}}/\lambda \approx 5\cdot 10^5$) Время затухания для спектральных линий, излучаемых газами, в действительности много больше, а сами спектральные линии много у́же, чем на кривой рис. 307.

Ширину спектральной линии принято характеризовать расстоянием между точками A и A', расположенными по разные стороны от центра линии, в которых интенсивность света равна половине интенсивности в центре линии. Это расстояние называется *полушириной спектральной линии*. Естественная полуширина спектральной линии, как это следует из формулы (89.5), равна

$$\Delta\omega = 1/\tau_{\text{изл}} = 2\gamma. \tag{89.6}$$

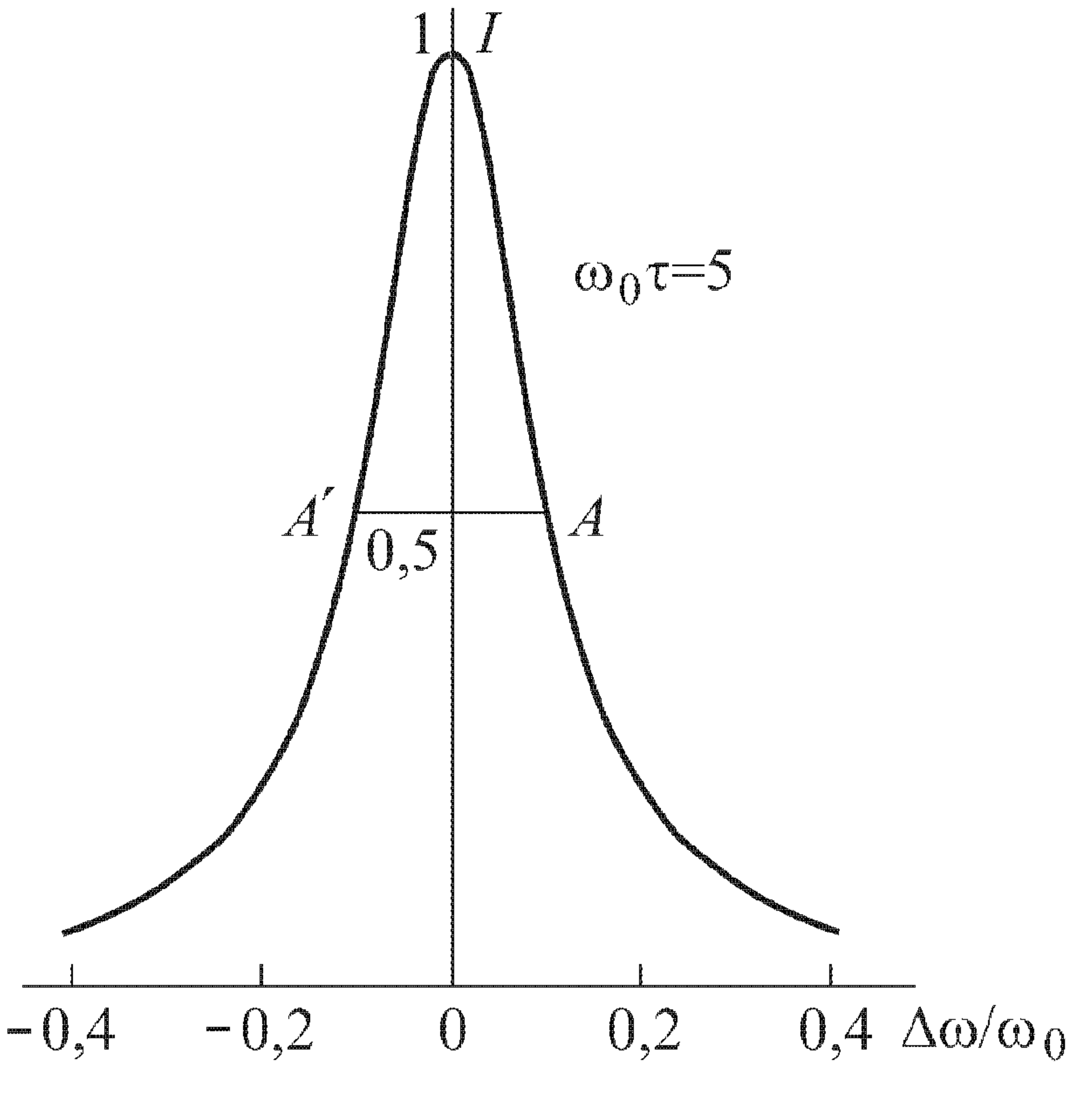


Рис. 307

В квантовой физике, где излучение связывают с переходом атома с более высокого энергетического уровня на более низкий (из возбужденного состояния в нормальное), величина $\tau_{\text{изл}}$ имеет смысл *времени жизни возбужденного состояния*. Здесь, как и в классической физике, формула (89.6) является следствием общего соотношения (29.8), согласно которому $\Delta\omega\cdot\tau_{\text{изл}} \sim 1$.

Естественное затухание свечения изолированных атомов и соответствующее ему уширение спектральных линий на опыте исследовалось В. Вином в 1919–1927 гг. В его опытах каналовые лучи, состоявшие из светящихся атомов, проходили через узкое отверстие ($0{,}1\times 3$ мм^2) в пространство, где с помощью мощных насосов поддерживался высокий вакуум ($<0{,}001$ мм рт. ст.). В этом пространстве атомы двигались без столкновений, но их свечение постепенно затухало по мере удаления от входного отверстия. По затуханию свечения можно было оценить время естественного затухания. Для этого надо было знать среднюю скорость движения атомов каналовых лучей. Она измерялась по доплеровскому смещению спектральных линий при наблюдении вдоль направления каналовых лучей и оказалась порядка $5 \cdot 10^7$ см/с (для атомов водорода). Из своих опытов Вин нашел для времени затухания τ около 10^{-8} с. Эта величина несколько менялась от одного вещества к другому и от одной спектральной линии к другой. Полученные результаты совпадали с предсказаниями теории, но только по порядку величины, а простая зависимость (89.4) времени затухания от длины волны ($\tau \sim \lambda^2$) не подтвердилась. Впрочем, полного количественного согласия и нельзя было ожидать от простой классической теории, основанной на модели гармонического осциллятора.

3. Все изложенное относится к излучению *изолированного атома*. В случае *среды*, состоящей из атомов, надо принять во внимание, что атом не только теряет энергию на излучение, но и получает энергию, излучаемую другими атомами. Если бы среда была *оптически однородна*, то, как показал Л.И. Мандельштам, эти два процесса в точности компенсировали бы друг друга. В отсутствие других причин затухания колебания атома были бы незатухающими. Плоская бегущая волна распространялась бы в среде без ослабления.

Дело меняется при нарушении оптической однородности среды. Однородность может нарушиться либо за счет *включения посторонних частиц* (например, пыли или капелек тумана в воздухе), либо за счет *тепловых флуктуаций*, возникающих в среде. В этом случае потеря энергии атомами на излучение при распространении волны будет приводить к ее ослаблению. Однако это не есть истинное поглощение света с переходом электромагнитной энергии волны в другие формы энергии (тепло), а *рассеяние света*, где такого превращения нет. *Истинное поглощение*, когда энергия волны превращается в другие формы (тепло), вызывается другими причинами. Кроме того, ослабление волны, обусловленное излучением, слишком мало по сравнению с тем ослаблением, которое во многих случаях наблюдается в действительности.

4. Одна из причин поглощения света состоит в том, что атомы, внутри которых происходят колебания, совершая тепловое движение, претерпевают *столкновения друг с другом*. При каждом столкновении резко и неправильно меняются амплитуды и фазы гармонических колебаний, происходит переход в тепло энергии регулярных колебаний, т.е. *поглощение света.* Исходя из этих представлений, Г.А. Лорентц развил теорию уширения спектральных линий, обусловленного столкновениями между атомами (молекулами) газа. Такое уширение называется *ударным уширением.* Лорентц показал, что в газах столкновения между молекулами при тепловом движении статистически приводят также к экспоненциальному закону затухания интенсивности волны и к форме спектральных линий такого же вида, что и при естественном затухании.

Мы не будем проводить детальные вычисления, как это сделал сам Лорентц. При таких вычислениях надо было бы вводить специальные предположения относительно характера нарушения регулярности амплитуд и фаз колебаний при столкновениях, так что результат оказался бы верным все равно с точностью до числового коэффициента. А в пределах такой точности его можно получить из общих соображений без всяких вычислений, если воспользоваться общей формулой (29.8).

Обозначим через $\tau_{\text{уд}}$ среднее время свободного пробега атома от одного столкновения до следующего. Можно считать, что между двумя последовательными столкновениями колебания внутри атома происходят гармонически, а при каждом столкновении обрываются. Затем они опять возобновляются, но уже с новыми значениями амплитуд и фаз, никак не связанными с амплитудами и фазами до столкновения. Но такие обрывающиеся во времени колебания при разложении в спектр будут заполнять интервал частот $\Delta\omega$, подчиняющийся соотношению $\Delta\omega \cdot \tau_{\text{уд}} \approx 1$. Отсюда следует, что ударное уширение спектральных линий определяется формулой

$$\Delta\omega \sim 1/\tau_{\text{уд}}. \tag{89.7}$$

Число столкновений, претерпеваемых атомом в единицу времени, определяется формулой кинетической теории газов $z = N\sigma v$, где N — число атомов в единице объема, σ — газокинетическое поперечное сечение атома, $v \approx \sqrt{kT/m}$ — средняя скорость теплового движения атома (см. т. II, § 86). Следовательно,

$$\tau_{\text{уд}} \approx \frac{1}{N\sigma v}. \tag{89.8}$$

Ударные затухание и уширение спектральных линий особенно существенны в *плотных газах* и при *высоких температурах.* Для уменьшения влияния столкновений надо уменьшать плотность газа. Вот почему в опытах Вина при изучении естественного затухания свечения атомов каналовые лучи направлялись в высокий вакуум. В обычных условиях столкновения значительно сильнее влияют на затухание волны, чем излучение. Однако формула (89.8) дает для времени затухания все же бóльшие, а следовательно, для уширения спектральных линий — меньшие значения, чем наблюдаются на опыте. Следовательно, должны существовать другие причины поглощения света и уширения спектральных линий.

Такими причинами могут быть воздействия на излучающий атом различных *силовых полей*, в первую очередь электрических полей, создаваемых окружающими атомами. В постоянных электрических полях спектральные линии претерпевают расщепление (эффект Штарка). В переменных полях изменение спектральных линий носит более сложный характер. Это — одна из основных причин уширения спектральных линий.

5. Другой причиной уширения спектральных линий является изменение частоты излучаемого света из-за *эффекта Доплера*. Пусть ω_0 — частота волны, излучаемой неподвижной молекулой (атомом). Если молекула равномерно движется, то частота волны, излучаемой ею в направлении оси X, в лабораторной системе отсчета изменяется и в нерелятивистском приближении определяется формулой

$$\omega = \omega_0(1 + v_x/c), \tag{89.9}$$

где v_x — скорость молекулы в направлении оси X.

Согласно закону распределения скоростей Максвелла, вероятность того, что x-составляющая скорости молекулы газа лежит в интервале $(v_x, v_x + dv_x)$, равна

$$dw = A_1 \exp\left(-\frac{mv_x^2}{2kT}\right) dv_x,$$

где m — масса молекулы, а A_1 — постоянная нормировки (см. т. II, § 72). Заменив в этом выражении переменную v_x на ω по формуле (89.9), получим

$$dw = A_1 \frac{c}{\omega_0} \exp\left[-\frac{mc^2}{2kT}\left(\frac{\omega - \omega_0}{\omega_0}\right)^2\right] d\omega.$$

Так как излучения различных молекул газа *некогерентны*, то эта величина определяет вероятность излучения в спектральный интервал $(\omega, \omega + d\omega)$. Интенсивность света $I(\omega)\,d\omega$, приходящаяся на этот интервал, будет пропорциональна dw. В результате получится

$$I(\omega) = I_0 \exp\left[-\frac{mc^2}{2kT}\left(\frac{\omega - \omega_0}{\omega_0}\right)^2\right], \tag{89.10}$$

где I_0 — новая постоянная. Очевидно, $I_0 = I(\omega_0)$, т.е. I_0 есть спектральная интенсивность в середине линии. По обе стороны от середины линии спектральная интенсивность убывает экспоненциально по закону (89.10). Полуширина спектральной линии, как нетрудно вычислить, определяется формулой

$$\Delta\omega = 2\sqrt{\frac{2kT}{mc^2}\ln 2}\,\omega_0 = 2\sqrt{\ln 2}\,\frac{v_m}{c}\,\omega_0 = 1{,}67\,\frac{v_m}{c}\,\omega_0, \tag{89.11}$$

где $v_m = \sqrt{2kT/m}$ — наиболее вероятная скорость молекулы.

Эффект Доплера является основной причиной уширения спектральных линий в разреженных газах при высоких температурах.

6. В основе теории дисперсии и абсорбции света, как она изложена в этой главе, лежат *линейные уравнения*. По этой причине показатели преломления n и затухания $\varkappa$ в этой теории *постоянны*, т.е. *не зависят от интенсивности света*. Амплитуда монохроматической волны при ее распространении убывает по закону $A = A_0 \exp(-2\pi\varkappa x/\lambda)$, а интенсивность — по закону

$$I = I_0 e^{-\alpha x}, \tag{89.12}$$

где $\alpha = 4\pi\varkappa/\lambda$ — *коэффициент затухания* (*поглощения*), зависящий, вообще говоря, от длины волны. Формула (89.12) выражает так называемый *закон Бугера* (1698–1758), постулированный им в 1729 г. Сущность этого закона состоит в том, что для монохроматического света коэффициент затухания (поглощения) α не зависит от интенсивности света. С.И. Вавилов, исследуя на опыте поглощение света в некоторых растворах, установил, что постоянство коэффициента поглощения α соблюдается с точностью не менее 5 % в крайне широком интервале изменения интенсивности света — от 20 до $2{,}5\cdot10^{-19}$ Вт/см^2, т.е. от интенсивности прямых солнечных лучей, сконцентрированных в фокусе линзы, до ничтожных интенсивностей, соответствующих порогу зрительного восприятия глаза.

Интерпретируем эти результаты с точки зрения *квантовых представлений* о поглощении света. При поглощении световая волна переводит молекулу (атом) с основного энергетического уровня W_0 на возбужденный уровень W_n, увеличивая тем самым запас поглощенной энергии. Способность молекулы поглощать свет в возбужденном состоянии иная, чем в основном. Если в опытах Вавилова коэффициент поглощения оставался одним и тем же при всех интенсивностях, то отсюда следует, что в этих опытах число возбужденных молекул было пренебрежимо мало по сравнению с числом невозбужденных, т.е. было относительно мало среднее время жизни молекулы в возбужденном состоянии. И действительно, для веществ, с которыми имел дело Вавилов в своих опытах, это время не превышало 10^{-8} с. Можно было бы ожидать, что при больших временах жизни и больших интенсивностях света число молекул в основном состоянии (которые практически только и поглощают свет) уменьшится и закон Бугера перестанет выполняться. Выбрав специально вещества с большими временами жизни в возбужденном состоянии, Вавилов наблюдал уменьшение коэффициента поглощения при увеличении интенсивности света. Это было первым примером, где наблюдалось *нарушение принципа суперпозиции в оптике.*

После открытия лазеров, сделавших возможным получение световых волн с напряженностями электрических полей, сравнимыми с внутримолекулярными и внутриатомными, класс подобных «нелинейных» оптических явлений быстро расширился. Их изучением занимается *нелинейная оптика* — самостоятельная обширная и быстро развивающаяся наука. Краткое представление об этих явлениях будет дано в последней (XI) главе этого тома.

§ 90. Двойное преломление в электрическом и магнитном полях

1. Электромагнитная природа света и вещества заставляет ожидать, что оптические свойства вещества должны изменяться при его внесении в электрическое и магнитное поле. Действительно, такие изменения были обнаружены на опыте. Сначала Фарадей в 1846 г. открыл магнитооптическое явление — *вращение плоскости поляризации света в магнитном поле.* Затем Керр (1824–1907) в 1875 г. открыл «электрооптическое явление», названное его именем. В 1877 г. тот же Керр нашел, что намагниченное железное зеркало иначе отражает поляризованный свет, чем ненамагниченное. В дальнейшем были открыты и другие *электрооптические* и *магнитооптические явления.*

Электрооптический эффект Керра состоит в том, что многие изотропные тела при введении в постоянное электрическое поле становятся оптически анизотропными. Они начинают вести себя подобно одноосным двупреломляющим кристаллам, оптическая ось которых параллельна приложенному электрическому полю.

Сам Керр наблюдал это явление сначала на стекле, вводимом в электрическое поле. В этом случае, а также в случае всех других твердых тел оптическая анизотропия может быть вторичным явлением, а именно может возникать в результате механических деформаций, производимых электрическим полем. Однако последующие исследования самого Керра и других ученых показали, что явление наблюдается также в жидкостях и даже газах, где механическая анизотропия исключена. Кроме того, «чистый эффект Керра» можно отделить от эффекта, вызванного механическими деформациями, исследуя явление в переменных полях: при внезапном включении или выключении электрического поля механические деформации возникают и исчезают не сразу, а с некоторой задержкой, тогда как явление Керра в чистом виде (наблюдаемое в жидкостях и газах) происходит *практически безынерционно* (см. п. 5).

Как и всякое двойное преломление, эффект Керра можно обнаружить, помещая между скрещенными николями N_1 и N_2 плоский конденсатор, заполненный исследуемым веществом (рис. 308, ср. с рис. 278). В отсутствие внешнего электрического поля свет через систему не проходит. При наложении электрического поля возникает *двойное преломление*, вследствие чего свет по выходе из конденсатора становится *поляризованным эллиптически* и частично проходит через николь N_2, — поле зрения просветляется. Вводя компенсатор K, можно измерить разность фаз между обоими лучами, возникшими при двойном преломлении, и таким образом найти разность $n_e - n_o$ между необыкновенным n_e и обыкновенным n_o показателями преломления исследуемого вещества в электрическом поле. Ввиду малых значений указанной разности ее измерение возможно только интерференционными методами, аналогичными изложенному.

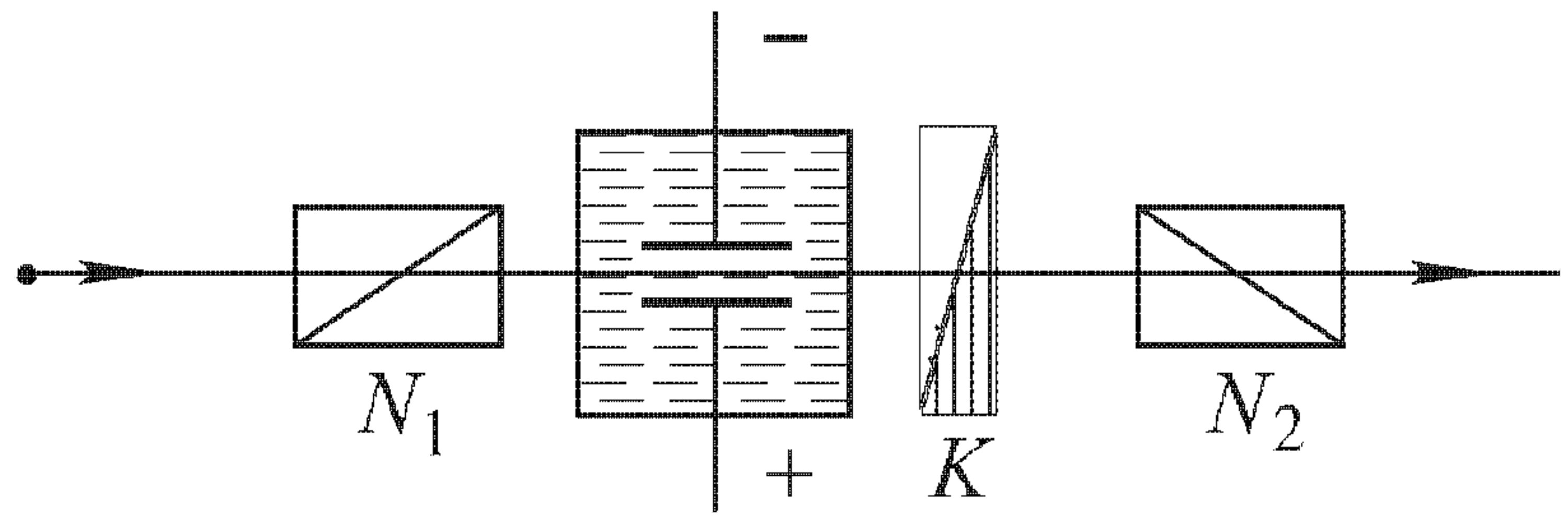


Рис. 308

Небольшим изменением описанного опыта можно для жидкостей найти обе разности $n_e - n$ и $n_o - n$. Для этого по предложению Л.И. Мандельштама можно воспользоваться, например, интерферометром Жамена. Исследуемая жидкость наливается в кювету, а конденсатор погружается в нее целиком (рис. 309). Один из интерферирующих лучей проходит внутри, а другой — снаружи конденсатора. Николь-поляризатор N устанавливается так, чтобы свет был поляризован либо в плоскости рисунка, либо перпендикулярно к ней. При наложении электрического поля показатель преломления n наружного луча не изменяется, тогда как луч, проходящий между пластинами конденсатора, становится либо необыкновенным, либо обыкновенным (в зависимости от установки николя). По смещению интерференционных полос можно вычислить $n_e - n$ и $n_o - n$.

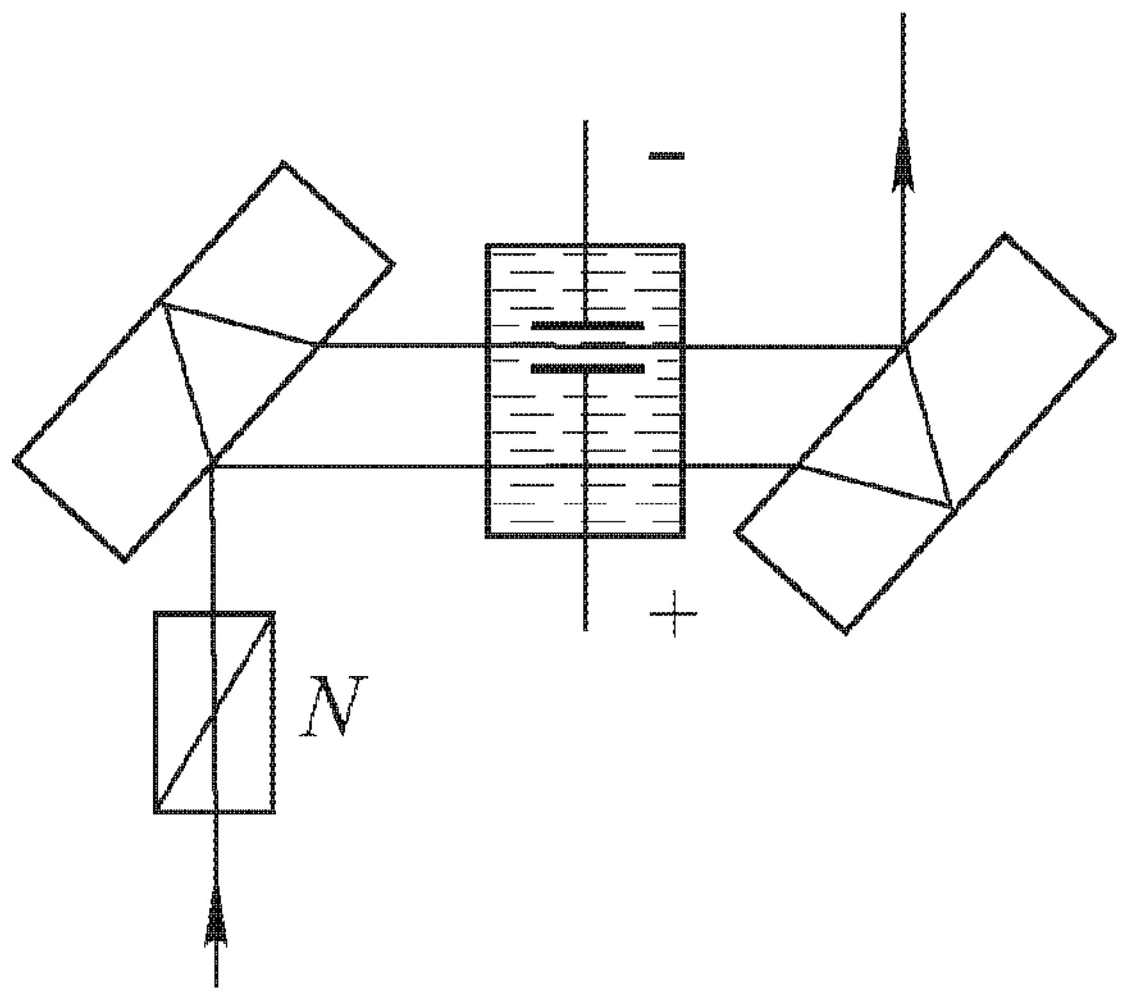


Рис. 309

2. В дальнейшем предполагается, что внешнее электрическое поле *однородно*, а луч распространяется перпендикулярно к нему. Внешнее поле обозначается через $\mathbf{E}_0$, в отличие от поля $\mathbf{E}$ самой волны. Очевидно, разность $n_e - n_o$ является функцией компонент поля E_{0x}, E_{0y}, E_{0z}. Если (свободная от поля) среда *изотропна*, то эта функция не может зависеть от направления вектора $\mathbf{E}_0$. В частности, она не должна изменяться при изменении знаков своих аргументов, т.е. быть *четной* по отношению к каждому из них. Поэтому ее разложение в степенной ряд должно содержать только *четные степени*. В слабых электрических полях, какими являются все поля, применяемые на опыте, можно сохранить только первые — *квадратичные* — члены этого разложения. В силу уже отмеченной изотропии среды они войдут только в виде квадратов E_{0x}^2, E_{0y}^2, E_{0z}^2 и притом с одним и тем же коэффициентом. Поэтому их можно объединить, написав

$$n_e - n_o = qE_0^2, \tag{90.1}$$

где коэффициент q зависит только от рода вещества и его состояния, а также от длины световой волны λ. Опыт действительно подтвердил формулу (90.1). Разность фаз между необыкновенным и обыкновенным лучами после прохождения через конденсатор будет

$$\varphi = \frac{2\pi}{\lambda}(n_e - n_o)l = 2\pi BlE^2, \tag{90.2}$$

где l — толщина проходимого слоя вещества, а $B \equiv q/\lambda$ — так называемая *постоянная Керра*. Она увеличивается при уменьшении длины волны (дисперсия) и сильно уменьшается с повышением температуры.

Таблица 9

Вещество	Температура, °C	Постоянная Керра B, СГСЭ-ед.	Вещество	Температура, °C	Постоянная Керра B, СГСЭ-ед.
Жидкости			*Газы*		
Нитробензол	20	$2{,}2 \cdot 10^{-5}$	Этиловый спирт	18	$9{,}2 \cdot 10^{-10}$
Нитротолуол	20	$1{,}2 \cdot 10^{-5}$	Ацетон	83	$5{,}4 \cdot 10^{-10}$
Хлорбензол	20	$1 \cdot 10^{-6}$	Сероуглерод	57	$3{,}6 \cdot 10^{-10}$
Вода	20	$4{,}7 \cdot 10^{-7}$	Этиловый эфир	63	$-0{,}66 \cdot 10^{-10}$
Сероуглерод	20	$3{,}2 \cdot 10^{-7}$	Этиленовый спирт	20	$-1{,}7 \cdot 10^{-10}$
Бензол	20	$6 \cdot 10^{-8}$			
Хлороформ	20	$-3{,}5 \cdot 10^{-8}$			

В табл. 9 приведены значения постоянной Керра для некоторых жидкостей и газов ($\lambda = 589$ нм, давление в случае газов 760 мм рт. ст.). Для большинства веществ $n_e > n_o$, т.е. постоянная Керра положительна. Это соответствует анизотропии *положительного кристалла.* Значительно реже встречаются случаи, когда $B < 0$ (этиловый эфир, многие масла и спирты). Максимальным значением B из всех исследованных веществ обладает нитробензол.

Чтобы составить представление о порядке величины эффекта Керра, приведем следующий пример. Пусть расстояние между пластинами «конденсатора Керра» равно 1 мм, а напряжение между ними 1500 В, так что $E_0 = 15000$ В/см $= 50$ СГСЭ-ед. Если конденсатор заполнен жидким нитробензолом, то при $l = 5$ см возникшая разность фаз будет $\varphi = \pi/2$. Такой конденсатор может служить пластинкой $\lambda/4$. Нетрудно обнаружить и значительно меньшие разности фаз. Поэтому эффект Керра в нитробензоле находит широкие технические применения.

3. Эффект Керра — явление более простое и теоретически лучше изучено, чем искусственная анизотропия при механических деформациях. Это и понятно. В последнем случае проявляется воздействие сложных молекулярных полей на молекулы тела, подвергшегося деформации. Оно плохо изучено и значительно сложнее соответствующего сравнительно простого воздействия постоянного однородного электрического поля. Кроме того, явление Керра удалось наблюдать в газах, для которых теория развита наиболее глубоко и подробно. Изучение явления Керра в газах совместно с рефракцией и деполяризацией рассеянного света позволяет определить тензор поляризуемости молекул.

Явление Керра объясняется анизотропией самих молекул. Количественная теория для газов была развита Ланжевеном (1872–1946) в 1910 г. В ней анизотропия молекулы характеризовалась *только тензором поляризуемости.* В отсутствие электрического поля анизотропные молекулы ориентированы в пространстве хаотически, так что среда в целом макроскопически изотропна. При наложении внешнего электрического поля молекулы преимущественно ориентируются осями наибольшей поляризуемости вдоль поля, вследствие чего среда становится анизотропной. В общем случае произвольного тензора поляризуемости вычисления очень громоздки. Однако сущность теории и ее основные результаты, по крайней мере качественно, можно передать, предполагая, что молекулы *полностью анизотропны,* что очень сильно упрощает вычисления. Это и делается ниже. *Полностью анизотропной* называется такая молекула, внутри которой электрические заряды могут смещаться только в определенном направлении, называемом *осью молекулы.* Моделью такой молекулы может служить палочка, вдоль которой и может происходить смещение зарядов[1]).

При проведении расчета (подчеркнем это еще раз) надо соблюдать осторожность, чтобы не спутать внешнее поле $\mathbf{E}_0$ с полем $\mathbf{E}$ электромагнитной волны, которое предполагается монохроматическим. Роль постоянного поля $\mathbf{E}_0$ сводится к созданию в среде определенного распределения молекул по их направлениям в пространстве. Переменное же поле $\mathbf{E}$ создает в ней переменную электрическую поляризацию $\mathbf{P}$, определяющую показатели преломления электромагнитной волны.

Переменный дипольный момент $\mathbf{p}$ полностью анизотропной молекулы создается только составляющей поля $\mathbf{E}$, параллельной ее оси. Он равен $\mathbf{p} = \beta E_s\mathbf{s} = \beta(\mathbf{Es})\mathbf{s}$, где $\mathbf{s}$ — единичный вектор в направлении оси молекулы, а β — ее поляризуемость в том же направлении. Очевидно, $\mathbf{p}$ можно представить в виде $\mathbf{p} = \beta E s_E \mathbf{s}$. Направление внешнего поля $\mathbf{E}_0$ примем за ось X, а направление распространения волны — за ось Z. Допустим сначала, что электрический вектор $\mathbf{E}$ волны направлен по оси X. Тогда $s_E = s_x$, и следовательно,

$$p_x = \beta E s_x^2, \quad p_y = \beta E s_x s_y, \quad p_z = \beta E s_x s_z.$$

Все направления, перпендикулярные к оси X, для распределения молекул равновероятны. Усредняя по этим направлениям, найдем: $\overline{s_x s_y} = \overline{s_x s_z} = 0$. Следовательно, для средних значений проекций дипольного момента получится

$$\overline{p}_x = \beta E\overline{s_x^2}, \quad \overline{p}_y = \overline{p}_z = 0,$$

а для вектора поляризации

$$P_x = N\beta E\overline{s_x^2}, \quad P_y = P_z = 0.$$

Вектор поляризации среды, таким образом, направлен по $\mathbf{E}$ и равен

$$\mathbf{P} = N\beta\mathbf{E}\overline{s_x^2}.$$

Соответствующая поляризуемость среды

$$\alpha_x = N\beta\overline{s_x^2}.$$

Показатель преломления будет *необыкновенным*. Для него и диэлектрической поляризуемости получаем

$$\varepsilon_x = n_e^2 = 1 + 4\pi N\beta\overline{s_x^2}.$$

---

[1]) Полностью анизотропных молекул, конечно, не существует. Поляризуемость реальной молекулы можно характеризовать *главными значениями* β_x, β_y, β_z тензора поляризуемости. Если одно из этих главных значений очень велико по сравнению с остальными двумя, то такая молекула может служить хорошим приближением к полной анизотропии.

Так как показатель преломления газа близок к единице, то

$$n_e \approx 1 + 2\pi N\beta\overline{s_x^2} = 1 + 2\pi N\beta\overline{\cos^2\vartheta}, \tag{90.3}$$

где ϑ — угол между осью молекулы и внешним полем $\mathbf{E}_0$ (т.е. осью X). Если распределение молекул изотропно (т.е. внешнего поля нет), то $\overline{s_x^2} = \overline{s_y^2} = \overline{s_z^2}$. Но $s_x^2 + s_y^2 + s_z^2 = 1$, а потому $\overline{s_x^2} = 1/3$. Значит, среда будет оптически изотропна, причем

$$n = 1 + \frac{2\pi}{3}N\beta. \tag{90.4}$$

Если поле $\mathbf{E}$ направлено вдоль оси Y, то $s_E = s_y$. В этом случае показатель преломления будет *обыкновенным*. Для него расчет, аналогичный приведенному выше, дает

$$n_o = 1 + 2\pi N\beta\overline{s_y^2}.$$

Чтобы вычислить $\overline{s_y^2}$, представим проекции единичного вектора $\mathbf{s}$ в полярной системе координат с полярной осью вдоль поля $\mathbf{E}$:

$$s_x = \cos\vartheta, \quad s_y = \sin\vartheta\cos\varphi, \quad s_z = \sin\vartheta\sin\varphi,$$

где φ — соответствующий азимут. Он с равной вероятностью может принимать все значения от 0 до 2π. Поэтому

$$\overline{s_y^2} = \overline{\sin^2\vartheta\cos^2\varphi} = \overline{\sin^2\vartheta}\,\overline{\cos^2\varphi} = \frac{1}{2}\overline{\sin^2\vartheta}.$$

Следовательно,

$$n_o = 1 + \pi N\beta\overline{\sin^2\vartheta}. \tag{90.5}$$

Прежде чем проводить дальнейшие вычисления, отметим одно следствие теории. Вычитая соотношение (90.4) из соотношений (90.3) и (90.5), получим

$$n_e - n = 2\left(\overline{\cos^2\vartheta} - \frac{1}{3}\right)\pi N\beta,$$

$$n_o - n = \left(\overline{\sin^2\vartheta} - \frac{2}{3}\right)\pi N\beta = \left(\frac{1}{3} - \overline{\cos^2\vartheta}\right)\pi N\beta.$$

Отсюда

$$\frac{n_e - n}{n_o - n} = -2. \tag{90.6}$$

Это соотношение, как показали тщательные измерения, хорошо выполняется для большинства веществ.

Найдем теперь постоянную Керра B. Из (90.3) и (90.5) получаем

$$n_e - n_o = 2\pi N\beta\left(\overline{\cos^2\vartheta} - \frac{1}{2}\overline{\sin^2\vartheta}\right) = 3\pi N\beta\left(\overline{\cos^2\vartheta} - \frac{1}{3}\right),$$

или на основании (90.4)

$$n_e - n_o = 9\frac{n-1}{2}\left(\overline{\cos^2\vartheta} - \frac{1}{3}\right). \tag{90.7}$$

Для дальнейших вычислений воспользуемся формулой Больцмана

$$dw = \mathrm{const}\cdot e^{-U/kT}\,d\Omega, \tag{90.8}$$

которая определяет вероятность dw направления оси молекулы в пределах телесного угла $d\Omega$. Потенциальная энергия молекулы в электрическом поле $\mathbf{E}_0$ равна $U = p^2/(2\beta) - pE_0\cos\vartheta$. Первый член есть энергия, затраченная на создание диполя с дипольным моментом $p = \beta E_0\cos\vartheta$, второй — потенциальная энергия этого диполя в электрическом поле $\mathbf{E}_0$. Сумма обоих членов $U = -\frac{1}{2}\beta E_0^2\cos^2\vartheta$. Поскольку она зависит только от угла ϑ, в качестве телесного угла $d\Omega$ удобно взять $2\pi\sin\vartheta\,d\vartheta$ и переписать формулу Больцмана в виде

$$dw = Ce^{-U/kT}\sin\vartheta\,d\vartheta, \tag{90.9}$$

где C — постоянная нормировки. Предполагая, что выполнено условие $|U/kT| \ll 1$, разложим экспоненциальную функцию в ряд и оборвем его на линейном члене:

$$e^{-U/kT} = 1 - \frac{U}{kT} = 1 + \frac{1}{2kT}\beta E_0^2\cos^2\vartheta.$$

Постоянную C определим из условия нормировки:

$$C\int_0^\pi \left(1 + \frac{1}{2kT}\beta E_0^2\cos^2\vartheta\right)\sin\vartheta\,d\vartheta = 2C\left(1 + \frac{1}{3}\frac{\beta E_0^2}{2kT}\right) = 1.$$

Следовательно,

$$dw = \frac{1}{2}\frac{1+\beta E_0^2\cos^2\vartheta/(2kT)}{1+\beta E_0^2/(6kT)}\sin\vartheta\,d\vartheta = \left[\frac{1}{2} + \frac{\beta E_0^2}{4kT}\left(\cos^2\vartheta - \frac{1}{3}\right)\right]\sin\vartheta\,d\vartheta,$$

$$\overline{\cos^2\vartheta} - \frac{1}{3} = \int\limits_{\vartheta=0}^{\vartheta=\pi}\left(\cos^2\vartheta - 1\right)dw = \frac{2}{45}\frac{\beta E_0^2}{kT}.$$

После подстановки в формулу (90.7) получим

$$n_e - n_o = \frac{n-1}{5}\frac{\beta E_0^2}{kT}, \tag{90.10}$$

а для постоянной Керра

$$B = \frac{n-1}{5\lambda kT}\beta. \tag{90.11}$$

4. Таким образом, разность $n_e - n_o$ пропорциональна *квадрату внешнего поля* E_0, как этого и следовало ожидать из соображений симметрии. Но в теории Ланжевена постоянная Керра всегда положительна и притом не только для полностью анизотропных молекул, но и для молекул с произвольным тензором поляризуемости. Борн (1882–1971) в 1916 г. устранил этот недостаток теории, распространив ее на *полярные молекулы* со значительными постоянными дипольными моментами $\mathbf{p}_0$, направления которых могут не совпадать с направлениями наибольшей поляризуемости молекул. Если $\mathbf{p}_0$ велико по сравнению с дипольным моментом молекулы, индуцированным внешним полем $\mathbf{E}_0$, то ориентация молекул в таком поле будет определяться преимущественно постоянными моментами. Постоянные моменты $\mathbf{p}_0$ будут стремиться ориентироваться вдоль поля $\mathbf{E}_0$, определяющего в среде направление оптической оси. Направление же наибольшей поляризуемости среды может составлять с ним заметный угол. Если эти два направления совпадают, то постоянная Керра B будет положительна. Если они взаимно перпендикулярны, то $B < 0$. В промежуточных случаях может получиться и тот и другой знак.

Конечно, наличие постоянного дипольного момента заметно усложняет вычисления, даже если молекулы полностью анизотропны. Ограничимся простейшим случаем, когда направления вектора $\mathbf{p}_0$ и наибольшей поляризуемости полностью анизотропной молекулы совпадают между собой. Тогда единственное отличие от случая, разобранного выше, состоит в том, что теперь потенциальная энергия молекулы в электрическом поле $\mathbf{E}_0$ будет $U = -(\mathbf{p}_0\mathbf{E}_0) = -p_0E_0\cos\vartheta$, и в разложении соответствующего экспоненциального множителя появится линейный член по E_0:

$$e^{-U/kT} = 1 + \frac{p_0E_0\cos\vartheta}{kT} + \frac{p_0^2E_0^2\cos^2\vartheta}{2k^2T^2}.$$

Но этот член не сказывается на вычислении $\overline{\cos^2\vartheta}$, так как интегралы от $\cos\vartheta\sin\vartheta$ и $\cos^3\vartheta\sin\vartheta$ обращаются в нуль. Таким образом, чтобы получить окончательный результат, достаточно в формуле (90.11) поляризуемость β заменить на p_0^2/kT. Это дает

$$B = \frac{n-1}{5\lambda}\left(\frac{p_0}{kT}\right)^2. \tag{90.12}$$

Формула (90.12) показывает, что и в случае полярных молекул разность $n_e - n_o$ пропорциональна квадрату поля E_0.

5. Поскольку процессы поляризации молекул и их ориентация в электрическом поле происходят за очень короткие времена, следует ожидать, что и время возникновения и уничтожения эффекта Керра при наложении и снятии электрического поля должно быть такого же порядка. Действительно, опыты Абрагама и Лемуана, начатые еще в 1899 г. и неоднократно продолженные другими исследователями, показали, что это время меньше 10^{-8} с, а в некоторых случаях даже меньше 10^{-9} с.

Точность всех этих опытов была недостаточна, чтобы высказать более определенные суждения. Это удалось только в результате развития лазерной техники, позволившей получать короткие импульсы (с длительностью порядка 10^{-12} с) мощного лазерного излучения (со средней напряженностью электрического поля в несколько десятков кВ/см).

Такие световые поля, хотя они и не являются статическими, способны вызывать двойное преломление в ячейках Керра. Принципиальная схема опыта приведена на рис. 310. Справа налево через ячейку Керра K проходит луч *1* голубого света, претерпевающий затем отражение от полупрозрачной пластинки S и попадающий в фотоумножитель ФЭУ. Если николи N_1 и N_2 скрещены, то свет в фотоумножитель не попадет. Однако если через ячейку Керра слева направо пропустить мощный импульс света от лазера L, то при прохождении через ячейку он вызовет в исследуемой жидкости двойное преломление луча *1*, и свет в фотоумножитель начнет поступать. Если на ту же ячейку наложить статическое электрическое поле и подобрать его напряженность так, чтобы возникла такая же разность фаз между необыкновенным и обыкновенным лучами, что и в предыдущем случае, то можно сравнить постоянные Керра в статическом и переменном полях лазерного излучения. Оказалось, что в недипольных жидкостях постоянные Керра в обоих случаях практически совпадают. В дипольных жидкостях при переходе от статического поля к переменному полю лазерного излучения постоянная Керра уменьшается. Так, в случае нитробензола она уменьшается приблизительно в 100 раз.

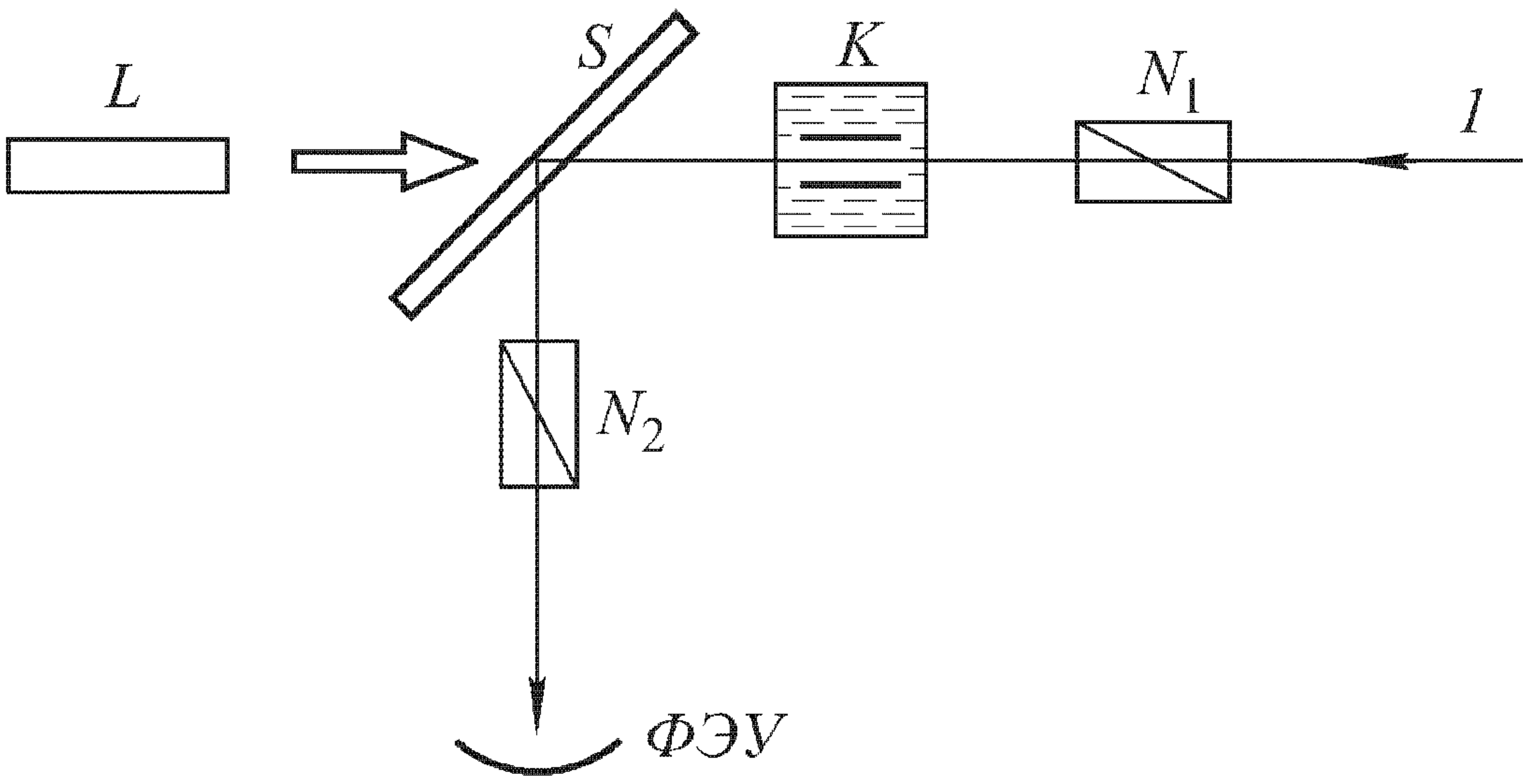


Рис. 310

Уже описанный опыт позволяет оценить длительность существования эффекта Керра. Более точные результаты дают измерения по схеме, приведенной на рис. 311. Мощный инфракрасный импульс от лазера с длительностью порядка 10^{-12} с и длиной волны $\lambda = 1060$ нм проходит через кристалл дигидрофосфата калия (KDP). Из-за нелинейного эффекта удвоения частоты (см. § 124) незначительная часть света превращается в зеленый свет с длиной волны $\lambda = 530$ нм. Зеркало S_1 пропускает инфракрасный луч и отражает зеленый. Зеркало S_2, наоборот, пропускает зеленый луч, но отражает инфракрасный. Оба луча проходят через ячейку Керра K с исследуемой жидкостью, расположенную между скрещенными николями N_1 и N_2. За николем N_2 поставлен светофильтр F, задерживающий инфракрасное излучение и пропускающий на фотоумножитель ФЭУ зеленое. Если после зеркала S_2 импульс зеленого излучения идет впереди мощного инфракрасного импульса, не перекрываясь с ним, то, разумеется, зеленый свет не сможет достигнуть фотоумножителя. Помещая на пути зеленого луча пластинку P, можно задержать зеленый импульс относительно инфракрасного и по толщине пластинки измерить время задержки τ. Располагая набором пластинок разной толщины, можно менять время τ. Опыты показали, что в простейшем случае интенсивность проходящего света, регистрируемая фотоумножителем, меняется экспоненциально по закону $I = I_0 \exp(-\tau/\tau_0)$. Постоянная τ_0 имеет смысл времени релаксации анизотропии жидкости, вызванной инфракрасным импульсом. Она и является мерой длительности эффекта Керра. Измерения дали для сероуглерода $\tau_0 = 2 \cdot 10^{-12}$ с, а для нитробензола $\tau_0 = 5 \cdot 10^{-11}$ с.

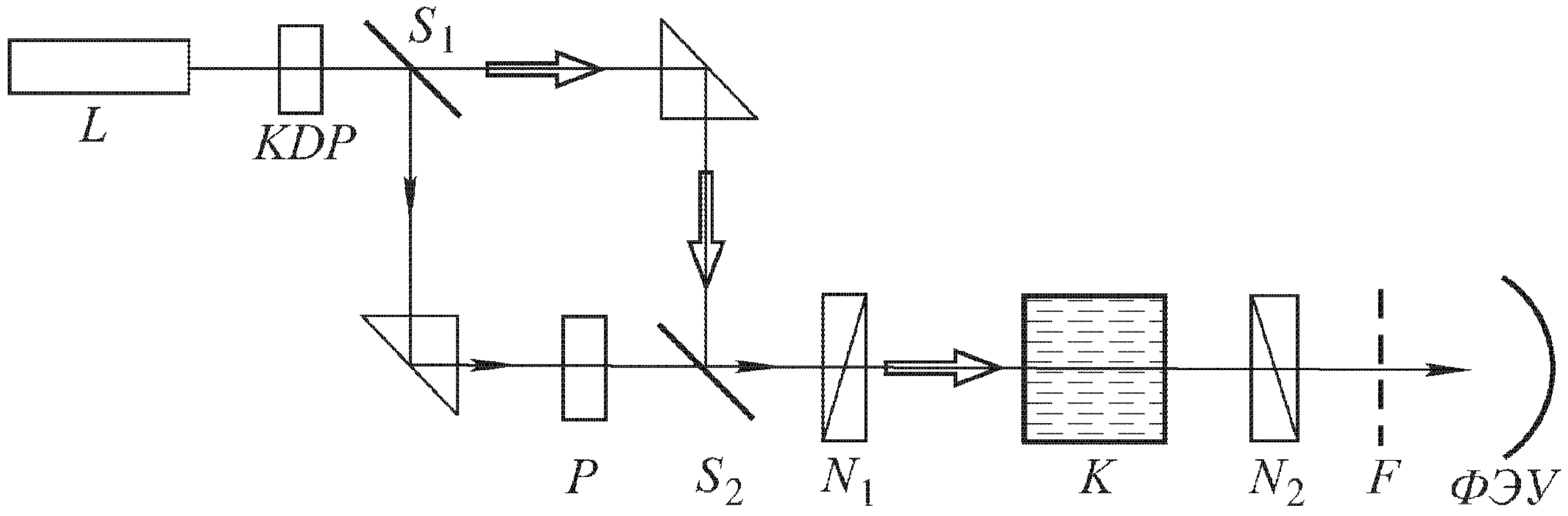


Рис. 311

6. Рассмотрим теперь механизм возникновения двойного преломления в переменных полях. Если молекулы полярные, то в одну половину периода их постоянные дипольные моменты стремятся ориентироваться в направлении внешнего электрического поля. В следующую половину периода, когда внешнее поле изменит направление на противоположное, возникнет такая же, но противоположно направленная ориентация. Эффекты ориентации, вызванные такими полями, вычитаются. В низкочастотных полях (в радиодиапазоне) в каждый момент времени устанавливается ориентация, соответствующая мгновенному значению электрического поля. В этом случае явление протекает так же, как в статических полях.

Однако в высокочастотных полях (с длинами волн короче примерно 1 см) ориентация постоянных дипольных моментов практически прекращается. В таких полях постоянные дипольные моменты $\mathbf{p}_0$ не могут играть роли в возникновении анизотропии, а с ней и двойного преломления среды. Ориентация обусловлена *только индуцированными дипольными моментами.* В одну половину периода, когда электрическое поле направлено в определенную сторону, индуцированные дипольные моменты создают моменты сил, стремящиеся приблизить оси наибольшей поляризуемости молекул к направлению электрического поля. В следующую половину периода направления всех моментов меняются на противоположные. Однако они также будут приближать оси наибольшей поляризуемости молекул к тому же направлению внешнего поля. Действительно, в отличие от вектора, у оси наибольшей поляризуемости нет одностороннего направления, *оба направления ее совершенно эквивалентны.* Не имеет никакого значения, каким концом эта ось приближается к направлению вектора $\mathbf{E}_0$. Таким образом, эффекты ориентации осей наибольшей поляризуемости в соседние половины периода, несмотря на противоположные направления векторов $\mathbf{E}_0$ в эти полупериоды, будут складываться, а не вычитаться. Хотя за каждый полупериод ориентация и ничтожна, благодаря такому сложению за время релаксации анизотропии τ_0 возникает достаточная ориентация, чтобы вызвать двойное преломление.

Таким образом, механизм возникновения двойного преломления в переменных полях высокой частоты — такой же, какой принимается в теории Ланжевена.

7. Благодаря чрезвычайной быстроте установления и исчезновения явления Керра оно нашло широкие научные и технические применения в качестве *быстродействующих затворов и модуляторов света.* Керровский модулятор света представляет собой ячейку Керра, конденсатор которой питается электрическим полем высокой частоты. Он позволяет осуществить громадное число (до 10^9) прерываний в секунду, недостижимое другими (например, механическими) средствами. Ячейка Керра, на которую подается кратковременный импульс электрического поля, может служить фотографическим затвором, время действия которого определяется длительностью этого импульса. Если в качестве электрического импульса взять мощный световой импульс от лазера, то время экспозиции можно довести до 10^{-12} с. Керровские затворы и модуляторы света применяются в лазерной технике для управления режимом работы лазеров.

8. Аналогом эффекта Керра является *эффект Коттона–Мутона*, открытый в 1910 г. Если молекулы среды анизотропны и обладают постоянными магнитными моментами, то они могут преимущественно ориентироваться постоянным магнитным полем. В достаточно сильных магнитных полях возникает анизотропия и связанное с ней двойное лучепреломление среды. Это и есть эффект Коттона–Мутона. Среда ведет себя подобно одноосному кристаллу, оптическая ось которого параллельна магнитному полю **B**.

Двойное преломление можно обнаружить и измерить так же, как это делается при изучении явления Керра. И законы обоих явлений совершенно аналогичны. Разность необыкновенного и обыкновенного показателей преломления определяется выражением $n_e - n_o = DB^2$, а соответствующая разность фаз

$$\varphi = \frac{2\pi}{\lambda}(n_e - n_o)l = 2\pi ClB^2, \tag{90.13}$$

где D и $C = D/\lambda$ — постоянные, зависящие от свойств среды и ее физического состояния. Для «постоянной Коттона–Мутона» в нитробензоле измерения дают $C = 2{,}23 \cdot 10^{-12}$ СГСМ. В магнитном поле $B = 20\,000$ Гс на пути $l = 10$ см возникает разность фаз $\varphi = 0{,}056$ рад, т.е. всего 3,2 градуса.

§ 91. Линейный электрооптический эффект Поккельса

1. Свободные колебания гармонического осциллятора описываются уравнением $\ddot{r} + 2\gamma\dot{r} + \omega_0^2 r = 0$. Они происходят около положения равновесия $r = 0$. Допустим теперь, что осциллятор находится в постоянном электрическом поле E_0. Тогда в отсутствие других внешних сил будет

$$\ddot{r} + 2\gamma\dot{r} + \omega_0^2 r = -\frac{e}{m}E_0.$$

Теперь положение равновесия сместится от начала координат на расстояние $r_0 = -eE_0/(m\omega_0^2)$. Обозначим через q расстояние колеблющейся частицы от нового положения равновесия. Тогда $r = r_0 + q$, и после подстановки в предыдущее уравнение имеем

$$\ddot{q} + 2\gamma\dot{q} + \omega_0^2 q = 0.$$

Отсюда видно, что в постоянном внешнем электрическом поле колебания осциллятора останутся гармоническими с *прежней частотой* ω_0, но они будут происходить около нового положения равновесия. Таким образом, постоянное электрическое поле не изменяет собственную частоту гармонического осциллятора, а только смещает положение равновесия, около которого совершаются свободные колебания.

В случае колебаний с большой амплитудой модель гармонического осциллятора может оказаться непригодной. В простейшем случае к квазиупругой силе $m\omega_0^2 r$ надо добавить член, пропорциональный квадрату смещения частицы из положения равновесия (начала координат). Свободные колебания такого *ангармонического осциллятора* описываются уравнением $\ddot{r} + 2\gamma\dot{r} + \omega_0^2 r + \beta r^2 = 0$, где β — постоянная. При наличии внешнего постоянного электрического поля E_0 уравнение колебаний переходит в

$$\ddot{r} + 2\gamma\dot{r} + \omega_0^2 r + \beta r^2 = -\frac{e}{m}E_0.$$

Теперь положение равновесия $r = r_0$ определится из уравнения

$$\omega_0^2 r_0 + \beta r_0^2 = -eE_0/m.$$

Из двух корней этого квадратного уравнения надо взять тот, который мало отличается от ранее найденного значения $r_0 = -eE_0/(m\omega_0^2)$ без учета ангармоничности (так как последняя предполагается малой). Пусть по-прежнему q означает отклонение колеблющейся частицы от нового положения равновесия, так что $r = r_0 + q$. Предполагая колебания малыми, пренебрежем квадратами q. Тогда

$$\ddot{q} + 2\gamma\dot{q} + (\omega_0^2 + 2\beta r_0)q = 0.$$

Отсюда видно, что во внешнем постоянном электрическом поле малые колебания ангармонического осциллятора в рассматриваемом приближении опять будут гармоническими. Однако *при наличии ангармоничности внешнее поле E_0 не только смещает положение равновесия, но и изменяет собственную частоту осциллятора.* Изменение квадрата собственной частоты осциллятора приближенно равно $\Delta\omega_0^2 = 2\beta r_0$, или в том же приближении

$$\Delta\omega_0^2 = -\frac{2e\beta}{m\omega_0^2}E_0. \tag{91.1}$$

2. Смещение собственных частот меняет кривую дисперсии, т.е. показатель преломления n среды. В простейшем случае, когда собственная частота ω_0 одна, величина n вдали от линии поглощения зависит только от разности $\omega^2 - \omega_0^2$, как это видно из формулы (84.9). Тогда изменение n в статическом электрическом поле E_0 определяется выражением

$$\Delta n = \frac{\partial n}{\partial\omega_0^2}\Delta\omega_0^2 = -\frac{\partial n}{\partial\omega_0^2}\frac{2e\beta}{m\omega_0^2}E_0.$$

Это выражение можно преобразовать, заметив, что $\partial f/\partial\omega_0^2 = -\partial f/\partial\omega^2$. Тогда

$$\Delta n = \frac{\partial n}{\partial\omega^2}\frac{2e\beta}{m\omega_0^2}E_0 = \frac{\partial n}{\partial\omega}\frac{e\beta}{m\omega\omega_0^2}E_0. \tag{91.2}$$

При фиксированном направлении внешнего поля $\mathbf{E}_0$ величина Δn зависит от направления распространения света. Это сказывается на двойном преломлении среды. *Изменение двойного преломления вещества из-за смещения собственной частоты во внешнем электрическом поле называется электрооптическим эффектом Поккельса.* В этом эффекте изменения показателей преломления пропорциональны *первой степени* внешнего поля E_0, в отличие от эффекта Керра, где они пропорциональны квадрату поля.

Эффект Поккельса может наблюдаться только в кристаллах, не обладающих центром симметрии. Дело в том, что он *линеен* относительно внешнего поля E_0. Поэтому при изменении направления поля $\mathbf{E}_0$ на противоположное должен меняться на противоположный и знак изменения Δn показателя преломления. Но в кристаллах с центром симметрии это невозможно, так как оба взаимно противоположных направления внешнего поля физически совершенно эквивалентны.

Из механизма явления ясно, что эффект Поккельса по крайней мере столь же *безынерционен*, что и эффект Керра. Поэтому он, наряду с эффектом Керра, нашел применение (например, в технике лазеров) в качестве *оптических затворов* и *высокочастотных модуляторов света.* Соответствующее устройство называется *ячейкой Поккельса.* Она представляет собой кристалл, помещаемый между двумя скрещенными николями. Такое устройство действует так же, как и ячейка Керра. Николи не пропускают свет, когда нет внешнего электрического поля, но при наложении такого поля пропускание появляется. Необходимо, чтобы кристалл до наложения внешнего электрического поля не давал двойного преломления. Этого можно достигнуть, если взять оптически одноосный кристалл, вырезанный перпендикулярно к оптической оси, а свет направить вдоль этой оси. Внешнее поле $\mathbf{E}_0$ может быть направлено либо перпендикулярно (*поперечный модулятор света*), либо параллельно распространению света (*продольный модулятор*).

§ 92. Эффект Зеемана

1. В 1896 г. Зееман (1865–1943) обнаружил, что спектральные линии определенным образом расщепляются, если источник света поместить в магнитное поле. В опыте Зеемана исследовалась очень узкая зелено-голубая линия кадмия и применялись магнитные поля с напряженностью 10 000–15 000 Гс. Г.А. Лоренц, развивавший в то время электронную теорию, сразу же объяснил *явление Зеемана* и тем самым придал дальнейшим исследованиям планомерный и целенаправленный характер.

Применяемая схема для наблюдения и исследования явления приведена на рис. 312. Источник света с линейчатым спектром (например, газоразрядная трубка или вакуумная дуга) помещается между полюсами электромагнита, создающего достаточно однородное магнитное поле. Исследуемый свет попадает на щель спектроскопа или спектрографа Sp с разрешающей силой около 100 000 или выше (дифракционную решетку или интерференционный спектральный аппарат). Николи N_1, N_2 и пластинка $\lambda/4$ служат для исследования поляризации излучаемого света. При фотографировании наблюдаемой картины применяются иногда многочасовые экспозиции. В течение всего этого времени должно быть обеспечено с достаточной точностью постоянство магнитного поля и температуры источника, чтобы картина оставалась неизменной во времени и можно было использовать спектральный аппарат высокой разрешающей силы.

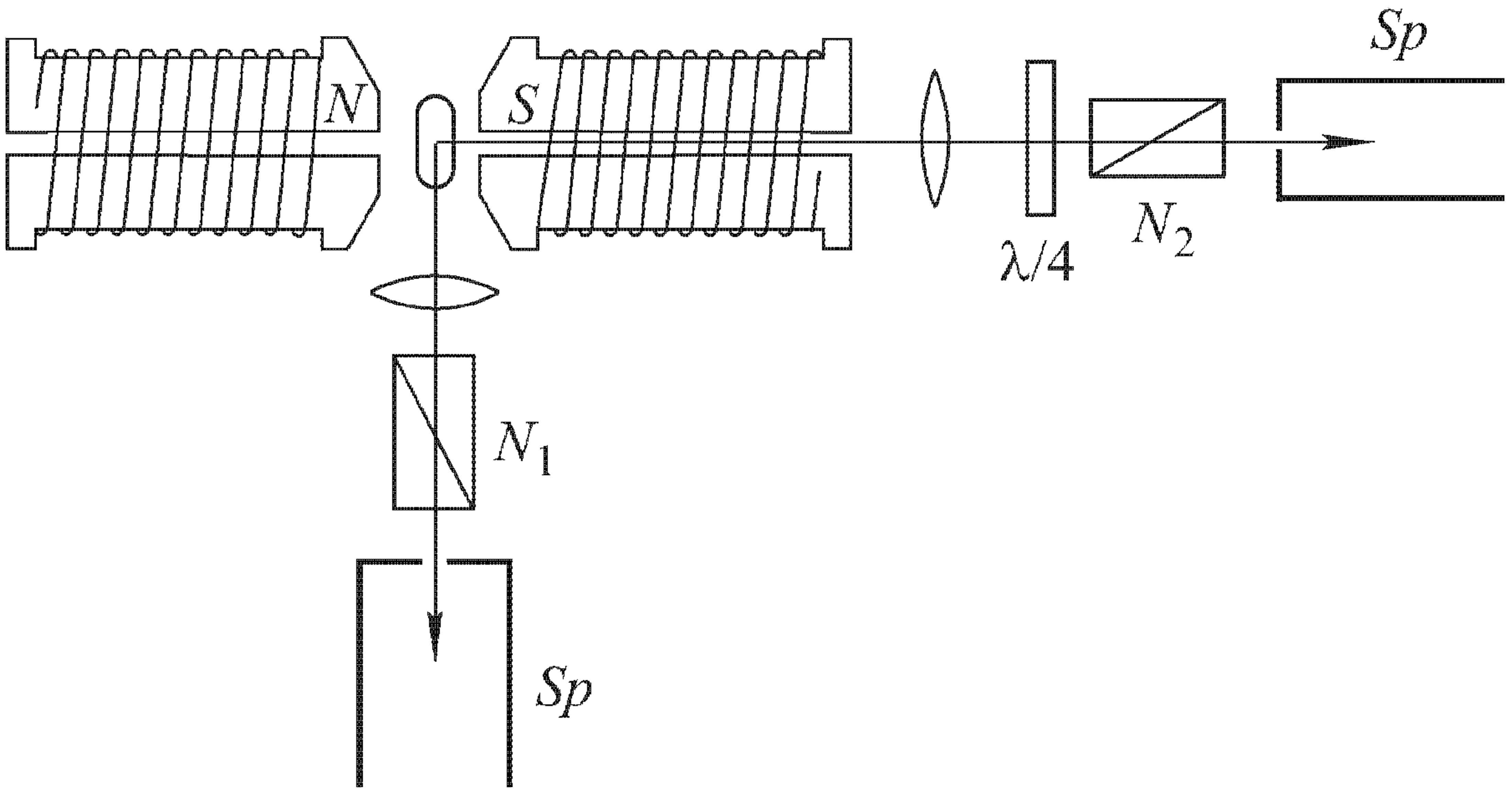


Рис. 312

В первых опытах Зееман обнаружил, что *при наблюдении поперек поля спектральная линия расщепляется на три линейно поляризованные компоненты.* Средняя компонента не смещена, крайние смещены в противоположные стороны на одинаковые расстояния (в шкале частот). Смещение пропорционально напряженности внешнего магнитного поля **B**. В средней компоненте электрический вектор направлен *параллельно магнитному полю* (такие компоненты называются *π-компонентами*), в крайних — *перпендикулярно к нему* (такие компоненты называются *σ-компонентами*). Интенсивность π-компоненты вдвое, а каждой из σ-компонент в четыре раза меньше интенсивности исходной линии.

При наблюдении *вдоль магнитного поля* получается *такое же смещение* (при одинаковой напряженности магнитного поля), что и в предыдущем случае, но *несмещенная компонента отсутствует.* Интенсивность каждой компоненты вдвое меньше интенсивности исходной спектральной линии. Обе компоненты *поляризованы по кругу* в противоположных направлениях (их принято называть также σ-компонентами). Если свет распространяется в направлении магнитного поля, то σ-компонента с меньшей частотой поляризована *по правому*, а с большей — *по левому* кругу. При изменении направления магнитного поля на противоположное меняется на противоположную и круговая поляризация обеих компонент.

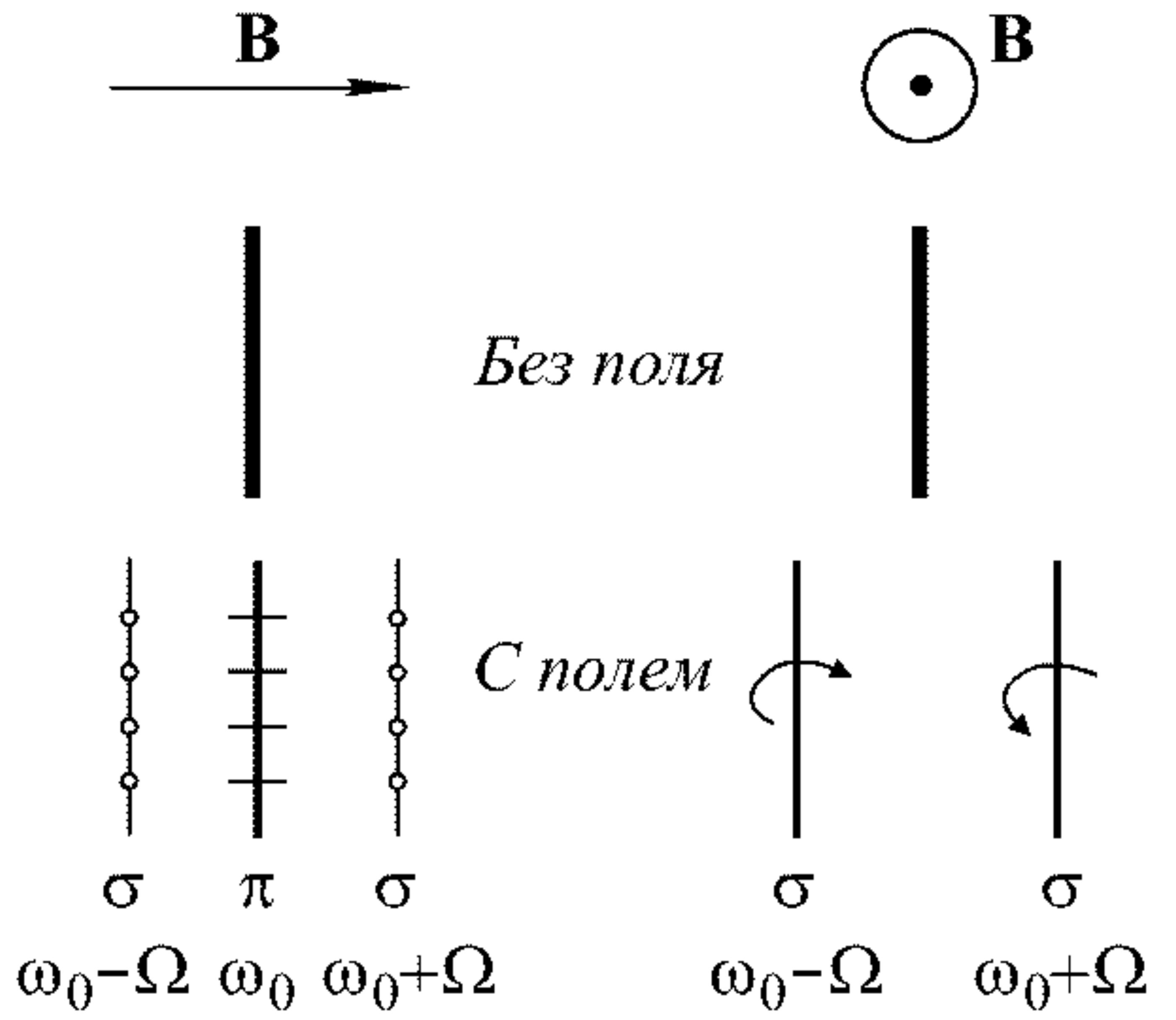


Рис. 313

Картина, наблюдаемая поперек и вдоль магнитного поля, представлена схематически на рис. 313. Предполагается, что в случае продольного эффекта свет распространяется вдоль магнитного поля, направленного к читателю. Относительные интенсивности линий показаны их толщиной, поляризация π-компоненты — штрихами, параллельными магнитному полю, а σ-компонент — кружочками.

2. Описанная картина расщепления спектральных линий объясняется *классической теорией Лоренца*. Как и классическая теория дисперсии, это есть *модельная теория*, в простейшей форме которой излучающими центрами являются *гармонические осцилляторы* в виде *квазиупруго связанных электронов*. В отсутствие внешнего магнитного поля уравнение движения такого электрона имеет вид $\ddot{\mathbf{r}} + \omega_0^2\mathbf{r} = 0$, где ω_0 — собственная частота электрона. При наличии постоянного магнитного поля на электрон действует еще сила Лоренца $-\frac{e}{c}\,[\dot{\mathbf{r}}\mathbf{B}]$ (заряд электрона обозначен через $-e$). Уравнение движения электрона принимает вид

$$\ddot{\mathbf{r}} + \omega_0^2\mathbf{r} = -\frac{e}{mc}\,[\dot{\mathbf{r}}\mathbf{B}],$$

где m — масса электрона. Введя ларморовскую частоту

$$\boldsymbol{\Omega} = \frac{e}{2mc}\mathbf{B}, \tag{92.1}$$

приведем его к виду

$$\ddot{\mathbf{r}} + 2[\dot{\mathbf{r}}\boldsymbol{\Omega}] + \omega_0^2\mathbf{r} = 0 \tag{92.2}$$

(см. т. III, § 86). Классическая теория сводится к решению этого уравнения. Для решения уравнения (92.2) перейдем к координатной форме. Направим ось Z прямоугольной системы координат вдоль магнитного поля $\mathbf{B}$. Тогда предыдущее уравнение сведется к системе трех скалярных уравнений

$$\begin{aligned} &\ddot{x} + 2\Omega\dot{y} + \omega_0^2 x = 0,\\ &\ddot{y} - 2\Omega\dot{x} + \omega_0^2 y = 0, \\ &\ddot{z} + \omega_0^2 z = 0. \end{aligned} \qquad (92.3)$$

Из последнего уравнения видно, что магнитное поле не влияет на движение электрона вдоль магнитного поля. Это и понятно, так как при таком движении не возникает силы, действующей со стороны магнитного поля. Интегрирование первых двух уравнений (92.3) удобно провести в *комплексной форме*. Объединим x и y в комплексную координату $\zeta = x + iy$. Она определяет положение электрона в координатной плоскости (X, Y) совершенно так же, как это делается с помощью двумерного вектора $\boldsymbol{\zeta}$ с составляющими x и y. Заметив, что $-i\dot{\zeta} = \dot{y} - i\dot{x}$, умножим второе уравнение (92.3) на i и сложим с первым. Тогда

$$\ddot{\zeta} - i \cdot 2\Omega\dot{\zeta} + \omega_0^2\zeta = 0.$$

Ищем решение этого уравнения в виде $\zeta = e^{i\omega t}$. Постоянная ω найдется из квадратного уравнения

$$-\omega^2 + 2\Omega\omega + \omega_0^2 = 0,$$

которое дает

$$\omega = \Omega \pm \sqrt{\omega_0^2 + \Omega^2}.$$

Даже в очень сильных магнитных полях квадратом ларморовской частоты можно пренебречь по сравнению с ω_0^2. Например, если $B = 10^4$ Гс, то формула (92.1) дает $\Omega \approx 10^{11}$ с$^{-1}$, тогда как для видимого света ($\lambda = 500$ нм) $\omega \sim 4 \cdot 10^{15}$ с$^{-1}$, а потому $(\Omega/\omega)^2 \sim 10^{-9}$. Максимальное магнитное поле, в котором измерялось зеемановское расщепление спектральных линий, получено в 1938 г. П.Л. Капицей (1894–1984). Оно было $3{,}2 \cdot 10^5$ Гс. Даже в этом случае $\Omega/\omega_0 \sim 1{,}4 \cdot 10^{-3}$, $(\Omega/\omega_0)^2 \sim 2 \cdot 10^{-6}$. Таким образом, с большой точностью $\omega = \pm\omega_0 + \Omega$. Чтобы не пользоваться отрицательными частотами, введем переобозначение, положив $\omega_1 = \omega_0 + \Omega$, $\omega_2 = \omega_0 - \Omega$. Тогда полученные два решения запишутся в виде

$$\zeta_1 = e^{i\omega_1 t}, \quad \zeta_2 = e^{-i\omega_2 t}.$$

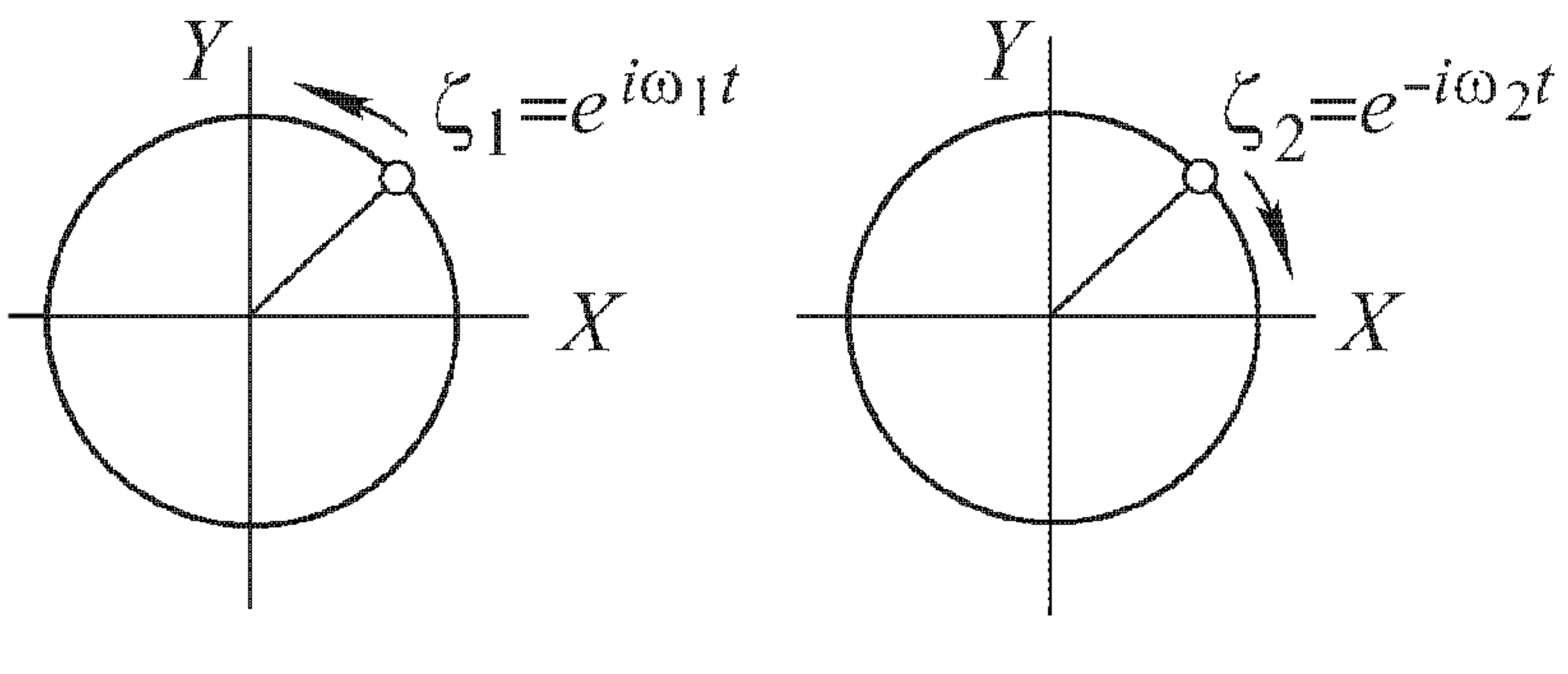


Рис. 314

Первое решение представляет *круговое движение*, в котором электрон вращается *против часовой стрелки* с угловой частотой ω_1, второе — также *круговое движение*, но *по часовой стрел-*

ке и с частотой ω_2 (рис. 314). Общее решение соответствует наложению таких двух вращений и представляется в виде $\zeta = C_1\zeta_1 + C_2\zeta_2$, где C_1 и C_2 — произвольные постоянные.

3. Чтобы нагляднее уяснить полученные результаты, разложим первоначальное движение электрона (т.е. движение в отсутствие магнитного поля) на два движения: на гармоническое колебание в направлении оси Z и на движение в плоскости XY. Второе движение в свою очередь разложим на два круговых вращения с одной и той же угловой частотой ω_0, но совершающиеся в противоположных направлениях. Тогда в постоянном магнитном поле колебание вдоль оси Z остается неизменным. Частоты же обоих круговых вращений изменяются на одну и ту же величину Ω: если вращение совершается против часовой стрелки, то частота увеличивается, а если по часовой стрелке, то уменьшается.

Для изменения частоты при вращении по кругу можно привести простое объяснение. Центростремительная сила, действующая на вращающийся электрон в отсутствие магнитного поля, равна $m\omega_0^2 r$. В магнитном поле к ней добавляется сила $\pm\frac{e}{c}\,vB = \pm\frac{e}{c}\,\omega r B$, так что новая центростремительная сила становится равной

$$\left(m\omega_0^2 \pm \frac{eB}{c}\omega\right) r = m(\omega_0^2 \pm 2\Omega\omega)r.$$

Выбор знака зависит от направления вращения. Приравнивая это выражение $m\omega^2 r$, приходим к уравнению $\omega^2 = \omega_0^2 \pm 2\Omega\omega$, из которого для положительных корней находим $\omega \approx \omega_0 \pm \Omega$. Это совпадает с результатами, полученными выше.

При включении магнитного поля кинетическая энергия вращения электрона изменяется. Возникает вопрос, как это может происходить, если сила, действующая со стороны магнитного поля, перпендикулярна к скорости электрона $\mathbf{v}$ и, следовательно, работы не совершает? Ответ состоит в том, что последнее утверждение относится к *постоянным магнитным полям*, которые только и учитываются уравнением (92.2). Но при включении магнитного поля оно *нарастает во времени* от нуля до максимального значения, а в дальнейшем вплоть до выключения остается постоянным. Во время же нарастания магнитного поля, согласно закону индукции Фарадея, возбуждается *вихревое электрическое поле*, которое и совершает работу над электроном, меняя его кинетическую энергию. Когда магнитное поле становится постоянным, электрическое поле пропадает и дальнейшее изменение кинетической энергии вращения электрона прекращается, пока не будет выключено магнитное поле. К этим установившимся вращениям и относятся движения, найденные выше. Подробное рассмотрение механизма изменения кинетической энергии вращения электрона было приведено в учении об электричестве (см. т. III, § 88).

4. Перейдем теперь к объяснению расщепления спектральных линий в магнитном поле. Колеблющийся электрон излучает электромагнитные волны. Излучение максимально в направлении, перпендикулярном к ускорению электрона, а в направлении ускорения отсутствует. Согласно классической теории, частота излучаемого света совпадает с частотой колебания электрона. Но последняя меняется при включении магнитного поля. Поэтому должна измениться и частота излучаемого света. При наблюдении вдоль магнитного поля колебание в том же направлении излучения не дает. Излучение создается *только круговыми вращениями электрона.* В результате наблюдаются две σ-компоненты с *круговой поляризацией* и частотами $\omega_0 + \Omega$ и $\omega_0 - \Omega$. Если свет идет в направлении вектора **B**, то поляризация первой линии будет *левой*, а второй — *правой*. При изменении направления магнитного поля на противоположное меняется на противоположную и круговая поляризация каждой линии. При наблюдении поперек магнитного поля **B** колебания электрона, параллельные **B**, дают максимум излучения. Им соответствует *несмещенная π-компонента*, в которой электрический вектор *параллелен* **B**. Оба круговых движения совершаются в плоскости, перпендикулярной к **B**. Разложим каждое из них на гармоническое колебание вдоль линии наблюдения и перпендикулярное к нему. Только колебания, перпендикулярные к линии наблюдения, сопровождаются излучением и дают две *σ-компоненты* с частотами $\omega+\Omega$ и $\omega-\Omega$, в которых электрические векторы *перпендикулярны к* **B**.

Таково объяснение расщепления спектральных линий, наблюдавшееся в первых опытах Зеемана. Если учесть, что в отсутствие магнитного поля все направления движения электрона равновероятны, то нетрудно объяснить и относительные интенсивности спектральных линий в этих опытах.

Как видно из численного примера, приведенного выше ($B = 10^4$ Гс), $\Omega/\omega_0 \approx 2\cdot 10^{-5}$. Для разрешения такого расщепления требуются спектральные приборы с разрешающей силой ω_0/Ω не менее $5\cdot 10^4$, т.е. дифракционные решетки или интерференционные спектроскопы. Но в опытах П.Л. Капицы ($B = 3{,}2\cdot 10^5$ Гс) были уже достаточны призменные спектроскопы.

Исследуя характер круговой поляризации линий в продольном эффекте Зеемана, можно определить *знак зарядов*, вызывающих этот эффект. Он оказался *отрицательным.* Измеряя же величину расщепления, можно определить удельный заряд e/m. Он оказался таким же, как и при измерениях по отклонениям катодных лучей в электрических и магнитных полях

($e/m = 1{,}759 \cdot 10^7$ СГСМ). Это не оставляет сомнений в том, что заряженные частицы, определяющие оптическое поведение атомов, действительно являются *электронами.*

5. Дальнейшие опыты показали, что явление Зеемана в том виде, в каком оно наблюдалось сначала и нашло объяснение в теории Лорентца — лорентцовский триплет, состоящий из одной π-компоненты и двух σ-компонент, а также дублет из двух σ-компонент, поляризованных по кругу, — наблюдается крайне редко. Такое расщепление называется *нормальным* или *простым эффектом Зеемана.* Простой эффект дают так называемые *синглетные*, т.е. одиночные, практически монохроматические спектральные линии. Подавляющее большинство спектральных линий являются *мультиплетами* (*дублетами*, *триплетами*, *квартетами* и т.д.), т.е. состоят из нескольких тесно расположенных спектральных линий.

Простейшим примером мультиплета (дублета) может служить двойная D-линия натрия. Она состоит из двух близко расположенных линий с длинами волн $\lambda_{D_1} = 589{,}5930$ нм и $\lambda_{D_2} = 588{,}96963$ нм, причем интенсивность линии D_2 вдвое больше интенсивности линии D_1.

Мультиплеты в магнитных полях дают значительно более сложную картину расщепления, чем расщепление в простом эффекте Зеемана. Так, линия D_1 натрия расщепляется на *четыре линии*: средние из них являются π-, а крайние — σ-компонентами. Линия же D_2 расщепляется на *шесть компонент*: две средние являются π-, а четыре крайние — σ-компонентами. Таким образом, весь дублет расщепляется на 10 линий. Наблюдаются и значительно более сложные картины расщепления мультиплетов. Такие расщепления называются *аномальным* или *сложным эффектом Зеемана.* Предпочтителен термин «сложный эффект», так как именно *сложный*, а не простой эффект *является правилом, а не исключением.*

Объяснение сложного эффекта Зеемана дала *квантовая теория*, да и то после того, как был открыт *спин* (т.е. собственный момент количества движения) и связанный с ним *магнитный момент электрона.* В случае синглетных спектральных линий квантовая теория приводит к тем же результатам, что и простая теория Лорентца. Мы вернемся ко всем этим вопросам в пятом томе нашего курса.

§ 93. Понятие об эффекте Штарка

Как было показано в § 91, однородное электрическое поле не меняет собственную частоту ω_0 гармонического осциллятора. Однако при учете ангармоничности такое изменение должно происходить. Если ограничиться линейными членами по внешнему полю E_0, то оно будет выражаться формулой (91.1), из которой следует $\Delta\omega_0 = -\beta e E_0/(m\omega_0^3)$. В этом приближении внешнее поле E_0 не вызывает расщепления спектральных линий, а только производит *небольшое смещение их в сторону*, пропорциональное напряженности поля E_0. При учете членов высших степеней можно ожидать расщепления спектральных линий с расстояниями между компонентами, пропорциональными *квадрату поля* E_0, а именно

$$\delta\omega_0 \sim [e/(m\omega_0^3)]E_0^2.$$

На возможность такого расщепления указывал Фохт (1850–1919). Однако из-за экспериментальных трудностей ему не удалось наблюдать ожидаемое явление. Трудность состояла в том, что для наблюдения эффекта требуются электрические поля E_0 порядка 100 000 В/см. В обычных же газоразрядных трубках, излучающих узкие спектральные линии (например, водородной серии Бальмера), такие сильные поля создавать и поддерживать невозможно из-за *сильной ионизации газа в трубках.*

Штарк (1874–1957) в 1913 г. преодолел эту трудность, создавая сильное электрическое поле за катодом в узком зазоре ($\sim$ 1 мм) между F и K (рис. 315), в котором непрерывной откачкой поддерживался высокий вакуум, так что в этом зазоре практически не происходило ионизации частиц из-за столкновений между ними. Возбуждение свечения атомов газа происходило в другой части трубки, как в обычной газоразрядной трубке. Светящиеся атомы (каналовые лучи) проникали через отверстия (каналы) в катоде и попадали в пространство между F и K, где и подвергались воздействию внешнего электрического поля.

Направление наблюдения

Рис. 315

С помощью такой установки Штарку удалось наблюдать расщепление спектральных линий в электрическом поле. Это явление было названо его именем. Оно оказалось совсем *не похожим* на то, что ожидал Фохт. Штарк исследовал расщепление спектральных линий водородной серии Бальмера. В видимой области спектра она содержит четыре линии: H_α (λ = 656,285 нм), H_β (λ = 486,132 нм), H_γ (λ = 434,046 нм), H_δ (λ = 410,173 нм). Расщепление *в водороде*, а также в случае всех *одноэлектронных ионов* оказалось пропорциональным *не второй степени элек-*

трического поля, а первой, т.е. значительно более сильным. По этой причине такое расщепление стали называть *линейным эффектом Штарка*.

Картина расщепления оказалась довольно сложной и различной для различных спектральных линий. Каждая спектральная линия водорода расщеплялась на несколько компонент. Для линии H_α было найдено расщепление на 15 компонент, для H_β на 20, для H_γ на 27 и для H_δ на 32 компоненты. В электрическом поле $E_0 = 104\,000$ В/см для расстояния между крайними компонентами Штарк нашел 2,3; 3,88; 5,88 и 7,5 нм для линий H_α, H_β, H_γ, H_δ соответственно. Эти расстояния довольно велики по сравнению с соответствующими расстояниями между крайними компонентами простого триплета Зеемана. (В поле $B = 104\,000$ Гс для указанных линий зеемановские расщепления составляли бы соответственно 0,42; 0,23; 0,18 и 0,16 нм.)

При наблюдении поперек поля $\mathbf{E}_0$ компоненты *поляризованы линейно*, причем часть из них является π-компонентами (у которых электрическое поле параллельно внешнему полю $\mathbf{E}_0$), а часть σ-компонентами (у которых электрическое поле перпендикулярно к внешнему полю $\mathbf{E}_0$). При наблюдении вдоль поля $\mathbf{E}_0$ π-компоненты не появляются, а σ-компоненты *не поляризованы*. Более сильные π-компоненты в общем случае лежат снаружи, а более сильные σ-компоненты — внутри. Расщепление и поляризация у водорода по обе стороны от первоначальной линии симметричны; у других атомов часто встречается несимметричное расположение. Расстояния (в шкале частот) компонент от первоначальной линии оказались кратными наименьшему возможному расстоянию между компонентами, причем опыты показали, что последнее одно и то же для всех линий бальмеровской серии водорода. В сильных полях (превосходящих примерно 100 000 В/см) были обнаружены отступления от линейной зависимости между расщеплением и напряженностью внешнего поля E_0. Линии атомов или ионов, содержащих *более одного электрона*, дают не линейный, а *квадратичный эффект Штарка*, в котором расщепление пропорционально *квадрату электрического поля* E_0.

Классическая теория не дает объяснения эффекта Штарка, как и всякого явления, связанного с излучением и поглощением спектральных линий. Его теория — существенно квантовая и не может быть изложена в этом томе.

§ 94. Вращение плоскости поляризации

1. Если линейно поляризованный свет проходит через плоскопараллельный слой вещества, то в некоторых случаях плоскость поляризации света оказывается повернутой относительно своего исходного положения. Это явление называется *вращением плоскости поляризации* или *оптической активностью.* Если вещество не находится во внешнем магнитном поле, то оптическая активность и вращение плоскости поляризации называются *естественными.* В противоположном случае говорят о *магнитном вращении плоскости поляризации* или *эффекте Фарадея.*

Естественная активность была открыта в 1811 г. Араго на пластинках кварца, вырезанных перпендикулярно к оптической оси. В 1815 г. Био подробно исследовал это явление, а также обнаруженное им вращение плоскости поляризации в растворах сахара. Затем естественное вращение плоскости поляризации было найдено у многих других тел. К концу прошлого века число известных естественно-активных веществ превышало 700. Теперь их известно гораздо больше, хотя для большинства веществ, где обнаружено это явление, оно выражено очень слабо.

Для наблюдения явления можно установить на оптической скамье два скрещенных николя. Такая система не пропускает свет. Однако, если между николями ввести пластинку кварца, вырезанную перпендикулярно к оптической оси, или слой какого-либо другого оптически активного вещества, то свет через систему будет проходить. Но его можно погасить вращением одного из николей. Отсюда следует, что после прохождения через активное вещество свет остается линейно поляризованным, но его плоскость поляризации оказывается повернутой. Для успеха опыта падающий свет, если он белый, необходимо монохроматизировать, пропустив его через светофильтр, так как *угол поворота плоскости поляризации зависит от длины волны.* Кварц — одноосный кристалл. В описанном опыте свет распространяется вдоль оптической оси, когда кварц ведет себя как изотропное тело, не давая обычного (линейного) двойного лучепреломления.

В зависимости от взятого вещества естественное вращение плоскости поляризации может происходить вправо или влево, причем эти два направления условились относить к наблюдателю, к которому свет приближается. В соответствии с этим различают *право-* и *левовращающие вещества.* Вращение вправо считается *положительным,* а влево — *отрицательным.*

2. Явление вращения плоскости поляризации указывает на определенную *дисимметрию*, свойственную оптически активным средам. Она выражается в том, что в таких средах направления вращения по и против часовой стрелки физически не эквивалентны. Поэтому в среде не может быть плоскости симметрии, проходящей через направление нормали к фронту волны. Иначе, как это следует из общих соображений симметрии, плоскость поляризации света не могла бы вращаться, если бы она совпадала с любой из плоскостей симметрии. В то же время естественно-активные среды, если они жидкие, *полностью изотропны*, т.е. все направления в них совершенно эквивалентны. Это проявляется, в частности, в том, что естественно-активная жидкость вращает плоскость поляризации *в одну и ту же сторону, независимо от направления распространения света.* Поэтому естественно-активную жидкость можно охарактеризовать как *дисимметрично-изотропную среду.* В кристаллах нет изотропии, но в одноосных кристаллах всякие два взаимно противоположные направления оптической оси также эквивалентны, по крайней мере в оптическом отношении.

Отмеченная дисимметрия напоминает дисимметрию винтовой спирали. Будем смотреть на один из торцов спирали. Пусть по спирали движется точка, вращаясь по часовой стрелке. Если при этом точка удаляется от нас, то спираль называется правой (в противоположном случае она называется левой). Если посмотреть на спираль с противоположного торца, то вращение той же точки будет происходить против часовой стрелки, но в этом случае точка будет приближаться к нам. Чтобы она удалялась, надо направление вращения изменить на противоположное. Таким образом, свойство спирали быть правой или левой не зависит от того, с какого торца на нее смотреть. Так и *свойство естественно-активной среды быть право- или левовращающей не зависит от того, в каком из двух прямо противоположных направлений распространяется свет.*

Таким образом, *если плоскость поляризации в естественно-активной среде вращается, например, вправо, то при изменении направления распространения света на противоположное она по-прежнему будет вращаться вправо. Однако направления вращения «вправо» и «влево» относятся к разным наблюдателям: к каждому из них свет должен приближаться. Объективно, независимо от выбора наблюдателя, вращения происходят в противоположные стороны, если лучи распространяются навстречу друг другу.* Если свет заставить пройти туда и обратно через естественно-активное вещество, отразив его от зеркала, то плоскость поляризации возвратится к своему исходному направлению.

3. Кварц встречается в природе в виде двух модификаций: *правовращающего* и *левовращающего* (короче — *правого* и *левого*). Это явление называется *энантиоморфизмом* и встречается у кристаллов, не содержащих центров и плоскостей симметрии. Обе энантиоморфные модификации кристалла отличаются друг от друга внешней формой и внутренней кристаллической структурой. По своей симметрии они отличаются друг от друга примерно так же, как правая спираль отличается от левой или правая рука от левой. Таким образом, обе модификации *не кон-*

груэнтны, т.е. правая не может быть наложена на левую и наоборот. Но зеркальное изображение одной из этих модификаций может быть совмещено с другой. По-видимому, все естественно-активные кристаллы существуют в двух энантиоморфных модификациях, хотя не во всех случаях известны обе модификации. Некоторые жидкости, например винная кислота, могут встречаться также в виде двух модификаций, вращающих плоскость поляризации в противоположных направлениях.

4. Био установил на опыте, что угол поворота χ плоскости поляризации пропорционален толщине l оптически активного вещества: $\chi = \alpha l$, где коэффициент α называется *вращением на единицу длины*. Он зависит от длины волны, природы вещества и температуры. Для кварца при температуре 20 °C и желтого света натрия ($\lambda = 589{,}3$ нм) $\alpha = \pm 21{,}728$, для хлорноватистокислого натрия ($NaClO_3$) $\alpha = 3{,}170$ угловых градуса на миллиметр. Для некоторых жидких кристаллов α может достигать 40 000 градусов на миллиметр. Для правых и левых модификаций кварца и всех остальных кристаллов значения вращения α *одинаковы по величине, но противоположны по знаку*. Вращение α увеличивается с уменьшением длины волны. Био нашел, что величина α обратно пропорциональна квадрату длины волны λ^2. Но такая зависимость грубо приближенна. В области прозрачности и малого поглощения хорошо согласуется с опытом формула Друде

$$\alpha = \sum_i \frac{B_i}{\lambda^2 - \lambda_i^2},$$

где B_i — постоянные, а λ_i — длины волн, соответствующие собственным частотам рассматриваемого вещества.

Оптическую активность жидкостей принято характеризовать *удельным вращением* $[\alpha]$. В случае химически чистых жидкостей это есть отношение вращения α к плотности жидкости ρ, т.е. $[\alpha] = \alpha/\rho$. В случае растворов речь идет обычно о растворах *активного вещества* в *неактивном растворителе*. В этом случае при определении удельного вращения $[\alpha]$ под ρ понимают *плотность активного вещества*, т.е. массу его в единице объема раствора. *Молекулярным вращением* жидкости или раствора называют произведение удельного вращения $[\alpha]$ на молекулярную массу активного вещества $[M] = M[\alpha]$. Зависимость вращения плоскости поляризации от концентрации активного вещества используется в *сахарометрии* для определения содержания сахара в растворе.

Вращение плоскости поляризации в одноосных кристаллах имеет место и в тех случаях, когда свет распространяется *под углом к оптической оси*. Но в этих случаях явление осложняется обычным двойным лучепреломлением. Еще больше усложнение в оптически двуосных кристаллах, так как вращение может быть разным вдоль каждой из оптических осей. Наконец, встречаются естественно-активные кристаллы кубической системы, не обнаруживающие обычного двойного лучепреломления (хлорноватистокислый натрий $NaClO_3$, бромноватистокислый натрий $NaBrO_3$). Для них вращение α не зависит от ориентации кристалла.

5. Согласно Френелю, вращение плоскости поляризации, какова бы ни была его физическая природа, есть проявление особого, так называемого *кругового двойного лучепреломления.* Однако две волны, которые могут распространяться в оптически активной среде с определенными, но различными скоростями, поляризованы не линейно, как при обычном (линейном) двойном лучепреломлении, а *по кругу: одна по левому, другая по правому.*

Для обоснования достаточности такого предположения допустим, что плоская волна с вращающейся плоскостью поляризации распространяется в положительном направлении оси Z. Так как волна поперечна, ее электрический вектор $\mathbf{E}$ будет содержать только две компоненты E_x и E_y. Если A — амплитуда колебаний, то $E_x = A\cos\chi\cos(\omega t - kz)$, $E_y = A\sin\chi\cos(\omega t - kz)$, где χ — угол между направлением электрического вектора $\mathbf{E}$ и осью X. Этот угол может зависеть от координаты z, так что можно положить $\chi = -\alpha z$, где α — введенный ранее коэффициент, определяющий вращение плоскости поляризации. Когда коэффициент α положителен, то при распространении волны угол χ уменьшается, т.е. вращение происходит *по часовой стрелке* (вправо). При отрицательном α плоскость поляризации вращается *против часовой стрелки* (влево). Подставив значение $\chi = -\alpha z$ в выражения для E_x и E_y и воспользовавшись известными формулами тригонометрии, представим их в виде

$$E_x = \frac{A}{2}\cos(\omega t - kz + \alpha z) + \frac{A}{2}\cos(\omega t - kz - \alpha z),$$

$$E_y = \frac{A}{2}\cos\left(\omega t - kz + \alpha z + \frac{\pi}{2}\right) + \frac{A}{2}\cos\left(\omega t - kz - \alpha z - \frac{\pi}{2}\right).$$

Разложим полученную волну на две: $\mathbf{E} = \mathbf{E}^{\text{п}} + \mathbf{E}^{\text{л}}$, где компоненты векторов $\mathbf{E}^{\text{п}}$ и $\mathbf{E}^{\text{л}}$ определяются выражениями

(94.1)

(94.2)

и введены обозначения

$$k^{\text{п}} = k - \alpha, \quad k^{\text{л}} = k + \alpha. \tag{94.3}$$

В волне (94.1) компонента E_y опережает по фазе компоненту E_x на $\pi/2$, а в волне (94.2) отстает на такую же величину. Обе волны поляризованы по кругу: в волне (94.1) электрический вектор вращается от положительного конца оси Y к положительному концу оси X, в волне (94.2) вращение происходит в противоположном направлении (см. рис. 314). Иначе говоря, волна (94.1) поляризована по правому, а волна (94.2) — по левому кругу. Таким образом, *кинематически волну с вращающейся плоскостью поляризации можно рассматривать как наложение двух волн, поляризованных по кругу: правому и левому.* Скорости этих волн определяются выражениями

$$v^{\text{п}} = \frac{\omega}{k - \alpha}, \quad v^{\text{л}} = \frac{\omega}{k + \alpha}, \tag{94.4}$$

а показатели преломления — выражениями

$$n^{\text{п}} = \frac{c}{v^{\text{п}}} = \frac{c}{\omega}(k - \alpha), \quad n^{\text{л}} = \frac{c}{v^{\text{л}}} = \frac{c}{\omega}(k + \alpha). \tag{94.5}$$

Отсюда

$$n^{\text{л}} - n^{\text{п}} = \frac{2c}{\omega}\alpha, \quad \alpha = \frac{\omega}{2c}(n^{\text{л}} - n^{\text{п}}). \tag{94.6}$$

Если $n^{\text{л}} - n^{\text{п}} > 0$, то $\alpha > 0$, т.е. вращение плоскости поляризации происходит *вправо.* Если же $n^{\text{л}} - n^{\text{п}} < 0$, то плоскость поляризации вращается *влево.* Иначе говоря, *плоскость поляризации вращается в ту же сторону, что и электрический вектор поляризованной по кругу волны с меньшим показателем преломления, т.е. с большей фазовой скоростью.*

6. Приведенное рассуждение отнюдь не доказывает, что каждая из поляризованных по кругу волн (94.1) и (94.2) может в отдельности существовать в среде. Мы исходили из опытного факта, что в оптически активной среде может реально существовать волна с вращающейся плоскостью поляризации. Такая волна, конечно, должна быть *решением системы фундаментальных уравнений Максвелла*, дополненной материальными уравнениями в оптически активной среде. Должна удовлетворять этой системе уравнений и суперпозиция поляризованных по кругу волн (94.1) и (94.2), так как мы доказали, что такая суперпозиция дает волну с вращающейся плоскостью поляризации. Но любое решение всякой системы уравнений можно представить (и притом бесконечным числом способов) в виде суммы нескольких слагаемых, которые *вовсе не обязательно должны быть решениями той же системы.* Поэтому из того факта, что в оптически активной среде возможна волна с вращающейся плоскостью поляризации, еще не следует, что в ней возможны и одиночные волны с круговой поляризацией. Однако *Френель предположил, что поляризованные по кругу волны со скоростями* (94.4) *действительно могут распространяться в оптически активной среде.*

Их можно назвать *нормальными волнами*, т.е. такими волнами, которые распространяются в среде с сохранением своей формы и характера поляризации. Это предположение есть *гипотеза*, которую теоретически Френель доказать не мог, так как для этого необходимо было бы располагать полной системой уравнений волновой теории света в оптически активных средах. Но Френель подтвердил свою гипотезу *экспериментально*.

Френель изготовил сложную призму, состоящую из трех кварцевых призм: двух призм ABM и DCM из правовращающего кварца и одной AMD из левовращающего с тупым углом $M = 152°$ (рис. 316). Оптические оси всех трех призм были параллельны основанию AD. Падающий луч, параллельный оптической оси, на грани AB не испытывал преломления, но разлагался внутри призмы ABM на два луча, поляризованные по кругу. Луч с правой поляризацией, как показывают формулы (94.5), имел меньший, а луч с левой поляризацией — больший показатель преломления ($n^{\text{п}} < n^{\text{л}}$). В левовращающей призме AMD соотношение между этими показателями преломления было обратным ($n^{\text{п}} > n^{\text{л}}$). Поэтому на грани AM лучи испытывали разное преломление: правый луч приближался к основанию AD, а левый удалялся. Угол расхождения между лучами еще больше увеличивался при преломлении на гранях DM и CD. В результате из призмы выходили два луча: один отклонялся вниз, а другой вверх, угол расхождения между которыми составлял около $4'$ (см. задачу к этому параграфу). Исследование с помощью параллелепипеда Френеля или пластинки $\lambda/4$ показало, что луч, отклонившийся вниз, был поляризован по правому, а вверх — по левому кругу.

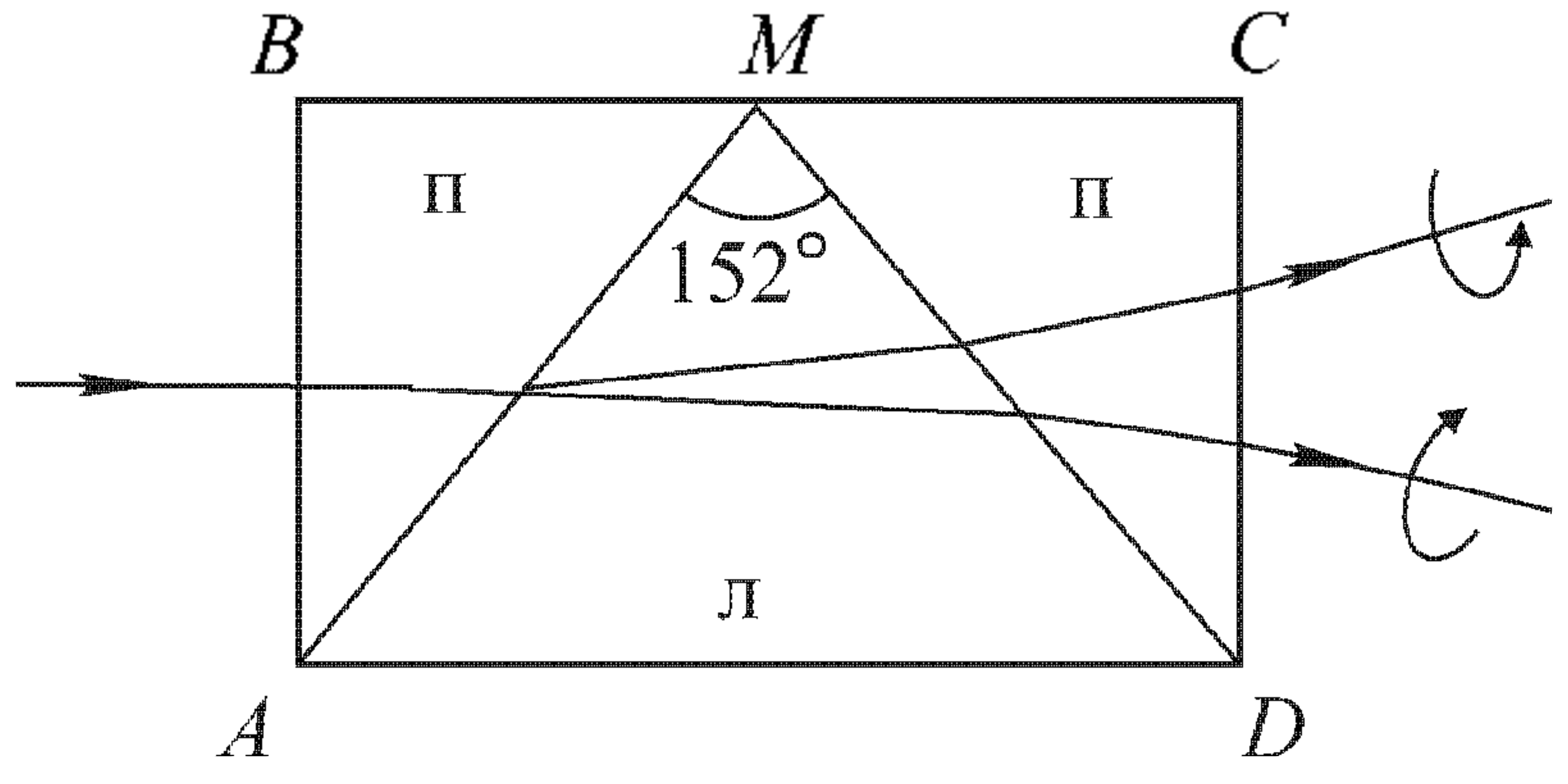


Рис. 316

Френель предлагал проверить свою гипотезу и для жидкостей путем комбинации большого ряда призм, попеременно наполненных жидкостями, вращающими плоскость поляризации вправо и влево. Позднее такой опыт был осуществлен и показал, что и в оптически активных жидкостях также могут существовать две волны с различной круговой поляризацией: правой и левой.

Существование двойного кругового лучепреломления в кварце можно обнаружить и с помощью одной призмы, вырезанной так, что ее оптическая ось перпендикулярна к плоскости, делящей пополам двугранный преломляющий угол. При наименьшем отклонении луч внутри призмы идет параллельно оптической оси, не испытывая обычного двойного преломления. Если падающий свет естественный или поляризован линейно и содержит только одну спектральную линию, то в минимуме отклонения при выходе из призмы спектральная линия расщепляется на две линии, из которых одна поляризована по правому, а другая по левому кругу.

Таким образом, Френель доказал экспериментально, что *при вступлении в оптически активную среду луч света испытывает двойное круговое лучепреломление: лучи, поляризованные по правому и левому кругу, идут внутри оптически активной среды с различными фазовыми скоростями. Если падающий свет был поляризован линейно, то при выходе из такой среды эти волны складываются снова в линейно поляризованную волну, но с повернутой плоскостью поляризации.* Тем самым задача объяснения вращения плоскости поляризации была сведена к задаче объяснения *кругового двойного лучепреломления.*

Поляризованные по кругу в противоположных направлениях волны в окрестности полос или линий поглощения могут отличаться не только скоростями распространения, но и коэффициентами поглощения. Тогда они выйдут с различными амплитудами. Если падающий свет был поляризован линейно, то выходящий будет поляризован *эллиптически.* Это явление называется *круговым дихроизмом.* Оно впервые (1847 г.) было замечено Гайдингером (1795–1871) на аметисте и подробно изучено (1895 г.) Коттоном (1869–1951) на растворах солей винной кислоты.

ЗАДАЧА

Пользуясь данными, приведенными в тексте, вычислить угол расхождения $\delta\varphi$ между лучами, поляризованными по правому и левому кругу, вышедшими из призмы $ABCD$ в опыте Френеля (рис. 316). Зная обыкновенный показатель преломления кварца ($n_o = 1{,}544$ для $\lambda = 589{,}3$ нм), найти угловое расхождение между теми же лучами $\delta_1\varphi$ после преломления на плоскости AM и $\delta_2\varphi$ после последующего преломления на плоскости MD.

Ответ. $\delta\varphi = \dfrac{4\lambda\alpha}{\pi}\operatorname{tg}\dfrac{M}{2} = 3{,}92'$; $\delta_1\varphi = \dfrac{\delta\varphi}{2n_o} = 1{,}27'$; $\delta_2\varphi = \dfrac{\delta\varphi}{n_o} = 2{,}54'$.

§ 95. Магнитное вращение плоскости поляризации

1. В 1846 г. Фарадей обнаружил, что естественно-неактивное вещество, помещенное в постоянное сильное магнитное поле, вращает плоскость поляризации, когда свет распространяется вдоль магнитного поля. Для наблюдения явления можно воспользоваться электромагнитом, вдоль оси которого просверлен цилиндрический канал, как это показано схематически на рис. 312. Электромагнит помещается между скрещенными николями, исследуемое вещество вводится между полюсами электромагнита. Если наблюдатель смотрит в направлении магнитного поля, то вращение вправо условились считать *положительным*, а влево — *отрицательным*. В первом случае вращение происходит в ту же сторону, куда течет ток по виткам электромагнита, во втором — в противоположную сторону. Фарадей наблюдал только положительное вращение. Таким вращением обладает большинство веществ. Из диамагнитных веществ единственным исключением является хлористый титан. Среди парамагнитных веществ встречается значительное число отрицательно вращающих, например соли железа и редких земель, $K_2Cr_2O_7$, $KCrO_4$ и др.

Опыты самого Фарадея, а затем более точные опыты Верде (1824–1866) показали, что *угол поворота плоскости поляризации пропорционален длине пути l света в веществе и напряженности внешнего магнитного поля, т.е. магнитной индукции B в веществе*:

$$\chi = RlB, \tag{95.1}$$

где коэффициент R называется *постоянной Верде* или *магнитной вращательной способностью*. Этот коэффициент зависит от рода вещества, его физического состояния и длины волны. В табл. 10 приведены характеристики магнитного вращения для некоторых веществ.

Таблица 10

Вещество	χ			
	$\lambda = 656$ нм	$\lambda = 589$ нм	$\lambda = 486$ нм	
Вода (25 °C)	0,0102′	0,0130′	0,0197′	На 1 см длины луча в поле 1 Гс
Сероуглерод (25 °C)	0,0319′	0,0415′	0,0667′	
Кварц	0,0136′	0,0166′	0,0250′	
Кислород	0,484′	0,559′	0,721′	При давл. 1 ат на 1 см длины луча в поле 10^5 Гс
Водород	0,430′	0,537′	0,805′	
Углекислота	0,691′	0,862′	0,286′	
Железо	217°	195°	145°	На 10^{-3} см в поле 15 000 Гс
Никель	92°	75°	64°	

В ферромагнитных металлах угол поворота χ не пропорционален B, но в сильных магнитных полях с возрастанием B возрастает очень медленно, приближаясь асимптотически к предельному значению. Согласно Кундту, угол χ пропорционален намагниченности магнетика $I = (\mu - 1)B/(4\pi\mu)$, где μ — магнитная проницаемость. Этот угол определяется выражением

$$\chi = KlI, \tag{95.2}$$

где K — так называемая *постоянная Кундта*. При $B = 15\,000$ Гс ферромагнетик можно считать намагниченным до насыщения.

2. Физическая природа магнитного вращения плоскости поляризации была выяснена после открытия явления Зеемана и объяснения его, данного Лорентцом. Допустим для простоты, что в отсутствие магнитного поля вещество испускает *только одну спектральную линию*, т.е. обладает *одной собственной* (*резонансной*) *частотой* ω_0. При внесении в магнитное поле $\mathbf{B}$ у осцилляторов вещества появляются две новые резонансные частоты $\omega_0 + \Omega$ и $\omega_0 - \Omega$ (где Ω — лармовская частота), соответствующие круговым вращениям электрона (см. § 92). Эти собственные частоты проявляются не только в *испускании* (*прямой эффект Зеемана*), но и в *поглощении света* (*обратный эффект Зеемана*). Прежняя резонансная частота ω_0 проявляется только при распространении света поперек магнитного поля и под углом к нему. Нормальные волны, которые могут распространяться вдоль магнитного поля, поляризованы по кругу. Когда направления распространения света и магнитного поля совпадают, большей собственной частоте $\omega_+ = \omega_0 + \Omega$ соответствует вращение по, а меньшей $\omega_- = \omega_0 - \Omega$ — против часовой стрелки, если смотреть в направлении магнитного поля (см. рис. 313).

Поскольку собственные частоты ω_+ и ω_- различны, скорости распространения обеих поляризованных по кругу волн также различны. Этим принципиально дано объяснение магнитного вращения плоскости поляризации, так как, согласно Френелю, при сложении таких поляризованных по кругу волн получается волна с вращающейся плоскостью поляризации. Угол поворота плоскости поляризации найдется из формулы (94.6). Она дает

$$\chi = \frac{\omega l}{2c}(n_- - n_+) = \frac{\pi l}{\lambda}(n_- - n_+), \tag{95.3}$$

где через n_+ и n_- обозначены показатели преломления соответственно для волн с собственными частотами ω_+ и ω_-. При выборе знака учтено, что положительные вращения плоскости поляризации в случае естественного и магнитного вращения в рассмотренном нами случае принято выбирать противоположно.

Формулу (95.3) удобно исследовать графически. На рис. 317 штриховыми линиями представлены графики для показателей преломления n_- и n_+ как функций частоты света ω. Их разность представлена сплошной линией. Как видно из рисунка, вне интервала (ω_-, ω_+) разность $n_- - n_+$, а с ней и вращение плоскости поляризации *положительны*, а внутри этого интервала — *отрицательны*.

Во всех случаях направления ларморовского вращения Ω, а следовательно, и вращения плоскости поляризации определяются *только направлением магнитного поля* **B** *и не зависят от направления распространения света.* При изменении на противоположное направления распространения света направление вращения плоскости поляризации *не изменяется*, в противоположность тому, что происходит при естественном вращении. Этим воспользовался Фарадей для усиления магнитного вращения плоскости поляризации. Он увеличил расстояние l, проходимое светом в веществе, заставив луч многократно отразиться от концов исследуемого образца (рис. 318). Для этого концы образца серебрились (за исключением мест входа и выхода светового луча).

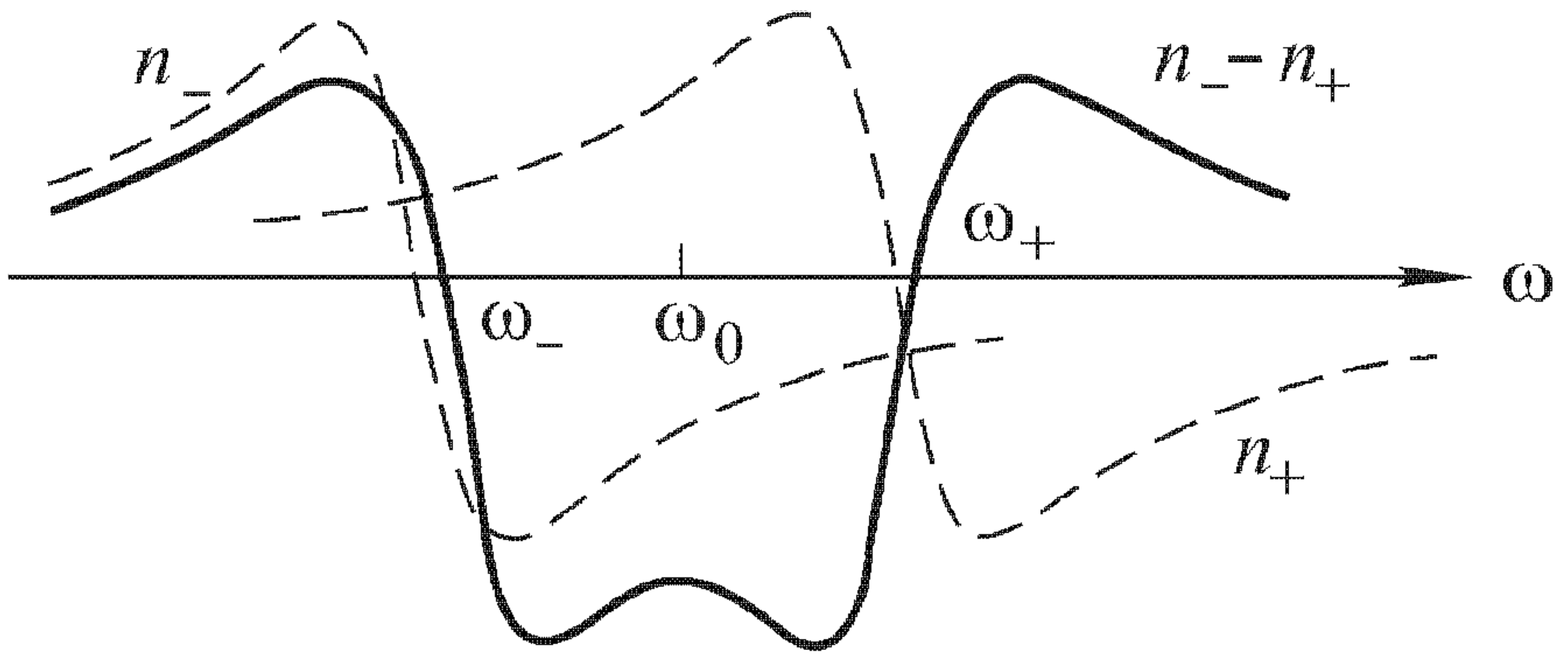


Рис. 317

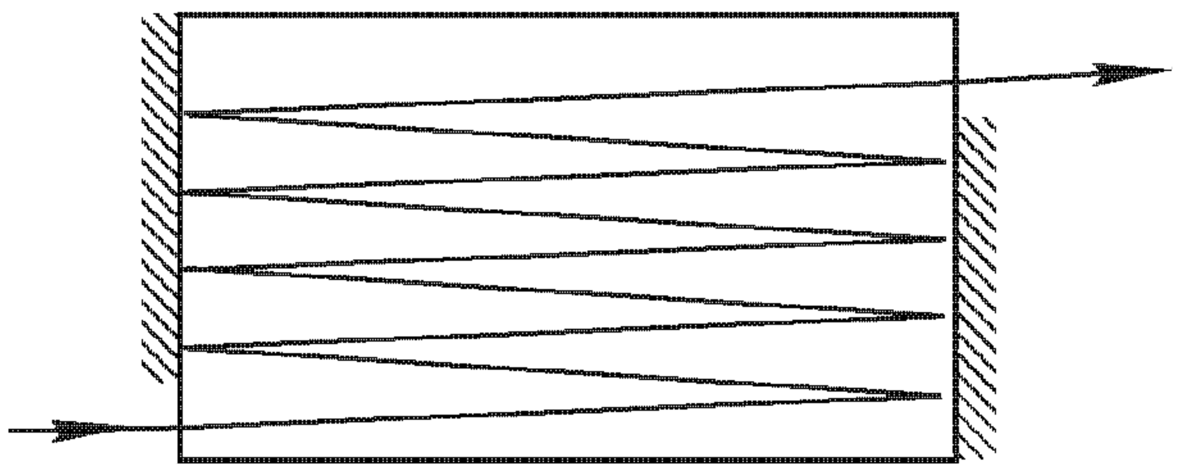


Рис. 318

3. Вне интервала (ω_-, ω_+) вдали от резонансных частот ω_- и ω_+ поглощением можно пренебречь и воспользоваться для показателей преломления n_- и n_+ формулой Зельмейера (84.9), заменив в ней ω_0 на ω_- и ω_+ для показателей преломления n_- и n_+ соответственно. Согласно этой формуле, n_- и n_+ являются *одной и той же функцией* аргументов $(\omega^2 - \omega_-^2)$ и $(\omega^2 - \omega_+^2)$. Такой же функцией, но от аргумента $(\omega^2 - \omega_0^2)$ выражается и показатель преломления n вещества в отсутствие магнитного поля. Следовательно,

$$n_- - n_+ = n(\omega^2 - \omega_-^2) - n(\omega^2 - \omega_+^2) =$$
$$= (\omega_+^2 - \omega_-^2)\frac{dn}{d\omega^2} = \frac{(\omega_+ + \omega_-)(\omega_+ - \omega_-)}{2\omega}\frac{dn}{d\omega}.$$

Вращение плоскости поляризации имеет заметную величину только не очень далеко от полосы поглощения, а потому можно принять, что $\omega_+ + \omega_- = 2\omega$. Кроме того, $\omega_+ - \omega_- = 2\Omega = eB/(mc)$. Следовательно,

$$n_- - n_+ = \frac{eB}{mc}\frac{dn}{d\omega} = -\frac{eB}{2\pi mc^2}\lambda^2\frac{dn}{d\lambda}.$$

Подстановка этого значения в (95.3) приводит к формуле (95.1), причем для постоянной Верде получается выражение

$$R = -\frac{e}{2mc^2}\lambda\frac{dn}{d\lambda}. \tag{95.4}$$

Нелишне напомнить, что здесь e означает заряд электрона *по абсолютной величине.* Формулу (95.4) обычно называют *формулой Беккереля* (1852–1908).

4. Перейдем теперь к исследованию магнитного вращения плоскости поляризации *в области полосы поглощения вещества.* Здесь показатель преломления, а также коэффициент поглощения меняются с частотой *быстро* и *аномально.* Поэтому магнитное вращение плоскости поляризации велико и носит также *аномальный характер.* Кроме того, внутри самой полосы поглощения между резонансными частотами ω_- и ω_+ знак вращения *отрицателен*, как это видно из рис. 317. Такое аномальное поведение плоскости поляризации называется *эффектом Макалюзо и Корбино.* Эти итальянские физики открыли его экспериментально в 1898 г., а затем Фохт дал теорию явления.

Макалюзо и Корбино показали, что пары натрия в магнитном поле вызывают очень сильное вращение плоскости поляризации в непосредственной близости к желтым линиям D_1 и D_2. Знак вращения одинаков по обе стороны этой двойной линии. Между полюсами электромагнита помещалось небольшое натровое пламя, через которое проходил пучок линейно поляризованного белого света в направлении магнитного поля. На пути лучей ставился николь-анализатор, который при выключенном магнитном поле полностью тушил свет. При включении тока в обмотке электромагнита поле зрения становилось освещенным ярким желтым светом: в спектроскопе были видны две желтые линии, симметрично расположенные по обе стороны линии D. Дело в том, что свет с такими длинами волн испытывал в магнитном поле поворот плоскости поляризации на 90° и потому проходил через скрещенные поляризатор и анализатор. Поворачивая николь-анализатор сначала в одном направлении, а затем в другом, можно было убедиться, что по обе стороны полосы поглощения знак вращения был один и тот же.

Вуд усовершенствовал установку Макалюзо и Корбино, произведя опыты с парами натрия, заключенными в стальной трубе. В его опытах плотность паров натрия можно было варьировать в значительных пределах. При сравнительно малой плотности наблюдалась такая же картина, как и в опытах Макалюзо и Корбино. Разница состояла в том, что в опытах Макалюзо и Корбино двойная D-линия натрия не разрешалась, а в опытах Вуда разрешалась. Желтые линии появлялись по обе стороны как компоненты D_1, так и компоненты D_2. При увеличении плотности паров натрия расстояние между наружными желтыми линиями увеличивалось и между ними появлялись новые линии, соответствующие вращению на 270°, 450°, 630° и т.д. Центр между линиями D_1 и D_2 последовательно становился то темным, то светлым. В некоторых случаях Вуд наблюдал до восьми таких перемен, что соответствует вращению на 1440°.

5. Изложенная теория основана на *классической теории*, т.е. *на простом эффекте Зеемана.* Полная теория должна учитывать *сложный эффект Зеемана*, т.е. быть *квантовой.* Кроме того, для объяснения отрицательного вращения плоскости поляризации необходимо учитывать *парамагнетизм атомов.* На этих вопросах мы останавливаться не можем.

ЗАДАЧИ

1. Между двумя николями N_1 и N_2 (рис. 319), главные сечения которых повернуты относительно друг друга на 45°, в продольном магнитном поле помещен слой вещества, вращающий плоскость поляризации. На какой угол этот слой должен вращать плоскость поляризации, чтобы система была *оптическим вентилем*, т.е. свет проходил бы через нее только в одном направлении (например, от тела A к телу B), а в обратном направлении не проходил?

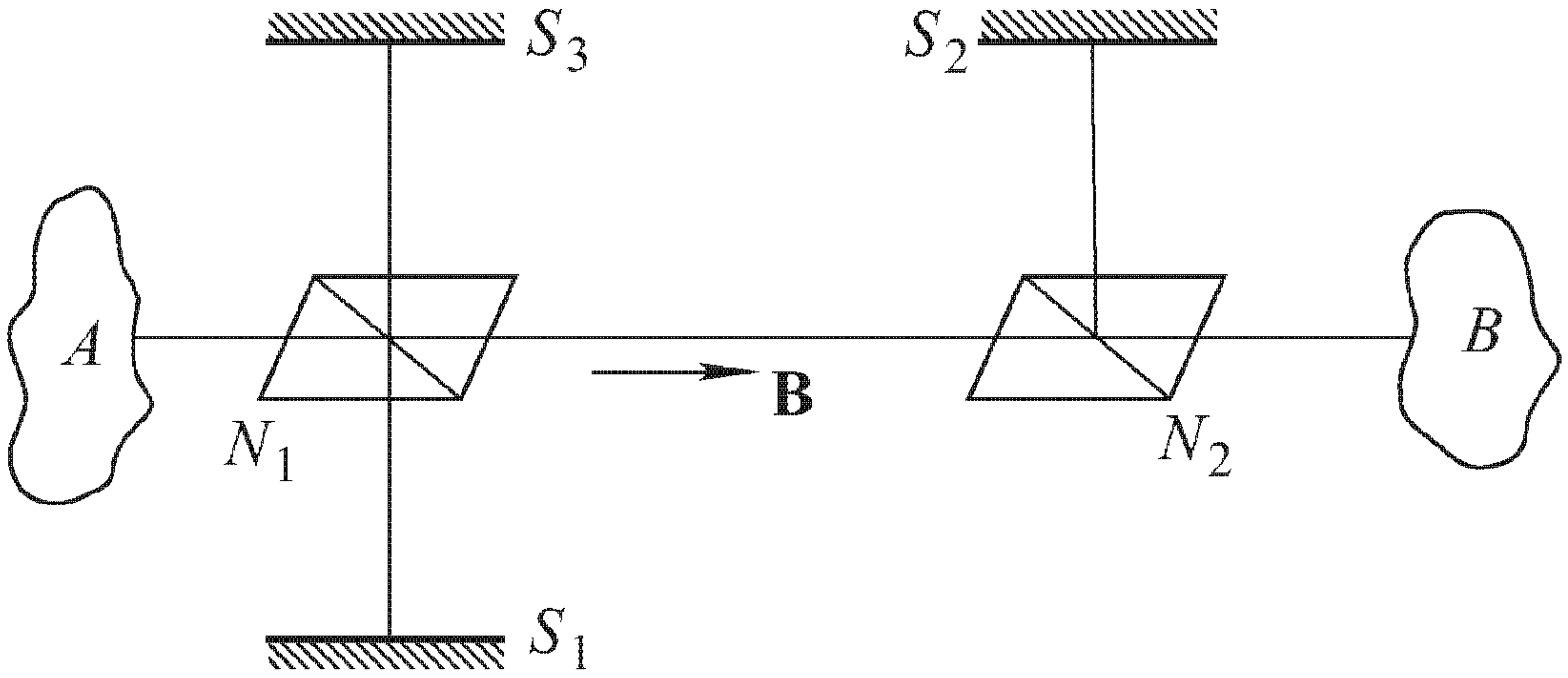


Рис. 319

Ответ. На угол $45° + 180° m$ (m — целое число).

2. Как будет вести себя оптический вентиль, описанный в предыдущей задаче, если направление магнитного поля изменить на противоположное, оставляя без изменения взаимное расположение николей и величину напряженности магнитного поля?

Ответ. Направление, в котором пропускается свет, изменится на противоположное.

3. Может ли установка, описанная в задаче 1, быть оптическим вентилем, если между николями поместить слой *естественно-активного вещества без магнитного поля*?

Ответ. Нет.

4. Разъяснить парадокс Вина, заключающийся в следующем. В адиабатической оболочке помещены два одинаково нагретых тела A и B (рис. 319), а между ними световой вентиль. Свет, излученный телом A, падает на николь N_1. Половина падающей энергии I, т.е. $I/2$, проходит через николь, а другая половина, испытав полное отражение в николе, возвращается зеркалом S_1 к телу A (применяя зеркало, можно не учитывать нагревание николя). Другая половина энергии $I/2$, пройдя через николь N_1, полностью пройдет через николь N_2 и поглотится телом B. Свет же от тела B не пройдет через вентиль. Половина излученной этим телом энергии I, равная $I/2$, испытает полное отражение в николе N_2 и после отражения от зеркала S_2 вернется к телу B. Другая половина энергии, пройдя через николь N_1 и вращающую среду, испытает поворот плоскости поляризации на 45° и поэтому не сможет пройти через николь N_1, а претерпит в нем полное отражение. Добавочным зеркалом S_3 эта часть энергии будет отражена обратно и после полного отражения в николе N_1 вернется к телу B. Таким образом, тело A получит энергию $I/2$, а тело B — энергию $3I/2$ и будет нагреваться в противоречии со вторым началом термодинамики.

Решение. Свет, излученный телом B (см. рис. 319) и возвращающийся к нему обратно после отражения от николя N_1, испытает при прохождении через вращающее вещество дополнительный поворот плоскости поляризации на 45° и поэтому не пройдет через николь N_2. Испытав в николе N_2 полное отражение, этот свет вернется (если поставить еще одно зеркало напротив S_2) к телу A. Приведенное решение парадокса Вина было дано Рэлеем.

§ 96. Временная и пространственная дисперсия. Теория естественной оптической активности

1. Многие оптические явления находят удовлетворительное объяснение в предположении, что связь между векторами $\mathbf{D}$ и $\mathbf{E}$ (а также между $\mathbf{B}$ и $\mathbf{H}$) *локальна во времени и пространстве*. Это значит, что вектор $\mathbf{D}$ в любой точке пространства $\mathbf{r}$ и в любой момент времени t определяется значением вектора $\mathbf{E}$ *в той же точке и в тот же момент времени*. (То же относится к векторам $\mathbf{B}$ и $\mathbf{H}$. В целях сокращения подобные замечания в дальнейшем подразумеваются, а все изложение ведется для векторов $\mathbf{D}$ и $\mathbf{E}$.) Однако для истолкования некоторых явлений предположения о локальной связи *недостаточно*. Пространственно-временную нелокальность можно разбить на *чисто пространственную* и *чисто временную*. Отвлечемся сначала от чисто пространственной нелокальности и учтем нелокальность временную. Среда во всем дальнейшем предполагается *однородной*.

2. Временная нелокальность проявляется в *оптической дисперсии среды*. Действительно, для истолкования дисперсии необходимо учитывать инерционные свойства электронов, атомных ядер и ионов вещества. А такая инерционность и приводит к нелокальной связи по времени между $\mathbf{D}$ и $\mathbf{E}$. Пренебрегая пространственной нелокальностью, будем рассуждать так, как если бы во всем пространстве электромагнитное поле было *однородно* и менялось *только во времени*. Действием магнитного поля будем пренебрегать. Рассмотрим сначала *импульсные воздействия* на среду. Допустим, что от момента $t = 0$ в течение времени dt среда подверглась воздействию электрического поля $\mathbf{E}$, а по истечении этого промежутка поле в среде снова обратилось в нуль. Такое воздействие может рассматриваться как *электрический толчок*, возбуждающий регулярные колебания электронов, атомных ядер и ионов среды. Эти колебания не прекращаются после прекращения действия поля, но продолжаются дальше в силу *конечности масс* колеблющихся частиц. В этом и проявляется временная нелокальность связи между отклонением частиц из положений равновесия и электрическим полем, а следовательно, между поляризацией $\mathbf{P}$ или индукцией $\mathbf{D}$ среды и полем $\mathbf{E}$.

В рамках *линейной электродинамики* поляризация среды $d\mathbf{P}$, вызванная электрическим толчком, пропорциональна $\mathbf{E}\,dt$. Предполагая, что среда *однородна* и *изотропна*, для вектора $d\mathbf{P}$ в момент времени t можно написать

$$d\mathbf{P}(t) = f(t)\mathbf{E}(0)\,dt, \tag{96.1}$$

где функция $f(t)$ зависит только от свойств среды и от времени t, которое прошло с момента действия толчка до момента наблюдения. Функция $f(t)$ должна обращаться в нуль при $t = 0$, так как из-за своей инерционности электроны, атомные ядра и ионы не могут мгновенно получить конечные смещения. Она должна обращаться в нуль и при $t = \infty$, так как все реальные среды *диссипативны*, так что всякое свободное колебание в них должно в конце концов затухнуть.

Если поле $\mathbf{E}$ действует в течение длительного промежутка времени, то этот промежуток можно разбить на бесконечно малые промежутки и таким путем свести воздействие электрического поля на среду к действию последовательных толчков. Вклад в поляризацию среды в момент времени t, внесенный более ранним электрическим толчком $\mathbf{E}(t')\,dt'$, будет $d\mathbf{P}(t) = f(t-t')\mathbf{E}(t')\,dt'$. В линейной электродинамике справедлив *принцип суперпозиции*, а потому полный вектор поляризации в момент времени t будет

$$\mathbf{P}(t) = \int\limits_{-\infty}^{t} f(t-t')\mathbf{E}(t')\,dt', \tag{96.2}$$

или, вводя новую переменную интегрирования $\vartheta = t - t'$,

$$\mathbf{P}(t) = \int\limits_{0}^{\infty} f(\vartheta)\mathbf{E}(t-\vartheta)\,d\vartheta. \tag{96.3}$$

Следовательно,

$$\mathbf{D}(t) = \mathbf{E}(t) + 4\pi \int\limits_0^\infty f(\vartheta)\mathbf{E}(t-\vartheta)\,d\vartheta. \tag{96.4}$$

Интегрирование производится по времени, *предшествующему* рассматриваемому моменту t. Этого требует *принцип причинности*. В своей нерелятивистской форме он означает, что *каждое событие определяется только прошедшими событиями, но не может зависеть от будущих*.

Если $\mathbf{E} = \mathbf{E}_0 \exp(i\omega t)$, то после подстановки этого выражения в (96.4) получится

$$\mathbf{D}(t) = \varepsilon(\omega)\mathbf{E}(t), \tag{96.5}$$

где

$$\varepsilon(\omega) = 1 + 4\pi \int\limits_0^\infty f(\vartheta)e^{-i\omega\vartheta}\,d\vartheta. \tag{96.6}$$

Таким образом, для монохроматического поля связь между $\mathbf{D}$ и $\mathbf{E}$ формально может быть записана в локальной форме с помощью диэлектрической проницаемости как функции частоты ω. Оптическая дисперсия, т.е. зависимость ε от ω, эквивалентна *нелокальной связи по времени между* $\mathbf{D}(t)$ и $\mathbf{E}(t)$. Поэтому такую дисперсию называют *временной* или *частотной дисперсией*, в отличие от *пространственной дисперсии*, о которой говорится ниже.

3. Учтем теперь *пространственную нелокальность*. Если электрическое поле *неоднородно*, то для нахождения индуцированного дипольного момента молекулы недостаточно знать вектор $\mathbf{E}$ в одной точке пространства, а требуется знание функции $\mathbf{E}(\mathbf{r})$ *во всем объеме, занимаемом молекулой*. Это эквивалентно знанию поля $\mathbf{E}$ и его пространственных производных всех порядков в какой-то одной точке внутри молекулы (которую условно можно назвать *центром молекулы*), так как тогда функция $\mathbf{E}(\mathbf{r})$ представится рядом Тэйлора по координатам x_1, x_2, x_3 вектора $\mathbf{r}$.

Легко оценить порядок последовательных членов этого ряда в монохроматическом поле световой волны. Производная $\partial E_l/\partial x_m$ будет порядка E/λ, а координата x_m — порядка размеров молекулы a, так что член $x_m \cdot \partial E_l/\partial x_m$ будет порядка $(a/\lambda)E$. Так же оцениваются и порядки членов, содержащих высшие производные. Если нулевой член ряда принять за единицу, то члены, содержащие первые, вторые и последующие производные, будут порядка a/λ, $(a/\lambda)^2$, $(a/\lambda)^3$ и т.д.

Как видно, в неоднородном поле связь между индуцированным дипольным моментом молекулы и электрическим полем $\mathbf{E}$

пространственно нелокальна. Это ведет к *пространственной нелокальности связи между* **P** *и* **E**, *а также между* **D** *и* **E**. При этом в слабых электрических полях, с которыми имеют дело линейная электродинамика и оптика, можно пренебречь влиянием квадратов и высших степеней поля **E** и его пространственных производных. В этом «линейном приближении» в монохроматическом поле световой волны можно написать

$$D_j = \varepsilon_{jl}E_l + \gamma_{jlm}\frac{\partial E_l}{\partial x_m} + \alpha_{jlmn}\frac{\partial^2 E_l}{\partial x_m\,\partial x_n} + \ldots \tag{96.7}$$

Для общности предполагается, что среда *анизотропна.* Ее оптические свойства характеризуются тензорами ε_{jl}, γ_{jlm}, ..., являющимися функциями частоты ω. В соответствии с общепринятой тензорной символикой по дважды встречающимся координатным индексам подразумевается суммирование. Как выяснено выше, если нулевой член ряда (96.7) принять за единицу, то последующие члены будут порядка a/λ, $(a/\lambda)^2$ и т.д.

В случае плоской монохроматической волны

$$\mathbf{E} = \mathbf{E}_0 \exp\left[i(\omega t - \mathbf{kr})\right]$$

дифференцирование по координате x_m сводится к умножению на $-ik_m$. Поэтому соотношение (96.7) можно записать в виде

$$D_j(\mathbf{r},t) = \varepsilon_{jl}(\omega,\mathbf{k})E_l(\mathbf{r},t), \tag{96.8}$$

где введено обозначение

$$\varepsilon_{jl}(\omega,\mathbf{k}) = \varepsilon_{jl}(\omega) - ik_m\gamma_{jlm} + (-ik_m)(-ik_n)\alpha_{jlmn} + \ldots \tag{96.9}$$

Таким образом, *в поле плоской монохроматической волны связь между* **D** *и* **E** *опять принимает формально локальный характер.* Однако тензор диэлектрической проницаемости $\varepsilon_{jl}(\omega,\mathbf{k})$ теперь зависит не только от ω, но и от **k**. Зависимость этого тензора от волнового вектора **k** называется *пространственной дисперсией.*

4. Ввиду малости параметра a/λ эффекты пространственной дисперсии в оптике малы и трудно наблюдаемы. Долгое время единственно известным из таких эффектов было *естественное вращение плоскости поляризации.* Оно наблюдается в таких средах, у которых тензор γ_{jlm} отличен от нуля.

Найдем вид тензора γ_{jlm} для дисимметрично изотропных сред — жидкостей и кристаллов кубической системы, у которых тензор $\varepsilon_{jl}(\omega)$ вырождается в скаляр. Отбросив в выражении (96.7) все члены, содержащие тензоры четвертого и высших порядков, напишем

$$D_j(\mathbf{r},t) = \varepsilon E_j(\mathbf{r},t) + \gamma_{jlm}\frac{\partial E_l}{\partial x_m}. \tag{96.10}$$

В развернутом виде для x-составляющей вектора $\mathbf{D}$ это соотношение гласит

$$D_x = \varepsilon E_x + \gamma_{xxx}\frac{\partial E_x}{\partial x} + \gamma_{xxy}\frac{\partial E_x}{\partial y} + \gamma_{xxz}\frac{\partial E_x}{\partial z} + \gamma_{xyx}\frac{\partial E_y}{\partial x} + \\ + \gamma_{xyy}\frac{\partial E_y}{\partial y} + \gamma_{xyz}\frac{\partial E_y}{\partial z} + \gamma_{xzx}\frac{\partial E_z}{\partial x} + \gamma_{xzy}\frac{\partial E_z}{\partial y} + \gamma_{xzz}\frac{\partial E_z}{\partial z}. \qquad (96.11)$$

Повернем теперь координатную систему вокруг оси X на угол 90° (рис. 320). Ввиду изотропии среды, все коэффициенты в предыдущем соотношении останутся неизменными. Однако y и E_y перейдут в $-z$ и $-E_z$, а z и E_z — в y и E_y. Остальные координаты и компоненты векторов $\mathbf{E}$ и $\mathbf{D}$ останутся неизменными. Произведя в (96.11) соответствующую замену, получим

$$D_x = \varepsilon E_x + \gamma_{xxx}\frac{\partial E_x}{\partial x} - \gamma_{xxy}\frac{\partial E_x}{\partial z} + \gamma_{xxz}\frac{\partial E_x}{\partial y} - \gamma_{xyx}\frac{\partial E_z}{\partial x} + \\ + \gamma_{xyy}\frac{\partial E_z}{\partial z} - \gamma_{xyz}\frac{\partial E_z}{\partial y} + \gamma_{xzx}\frac{\partial E_y}{\partial x} - \gamma_{xzy}\frac{\partial E_y}{\partial z} + \gamma_{xzz}\frac{\partial E_y}{\partial y}. \qquad (96.11a)$$

Но в силу определения тензора γ_{jlm} в повернутой системе координат соотношение между D_x и компонентами вектора $\mathbf{E}$ можно записать в прежней форме (96.11). Сравнение обоих соотношений дает $\gamma_{xxy} = \gamma_{xxz}$, $\gamma_{xxz} = -\gamma_{xxy}$, откуда $\gamma_{xxy} = \gamma_{xxz} = 0$ и т.д. Кроме того, $\gamma_{xyz} = -\gamma_{xzy}$ и т.д.

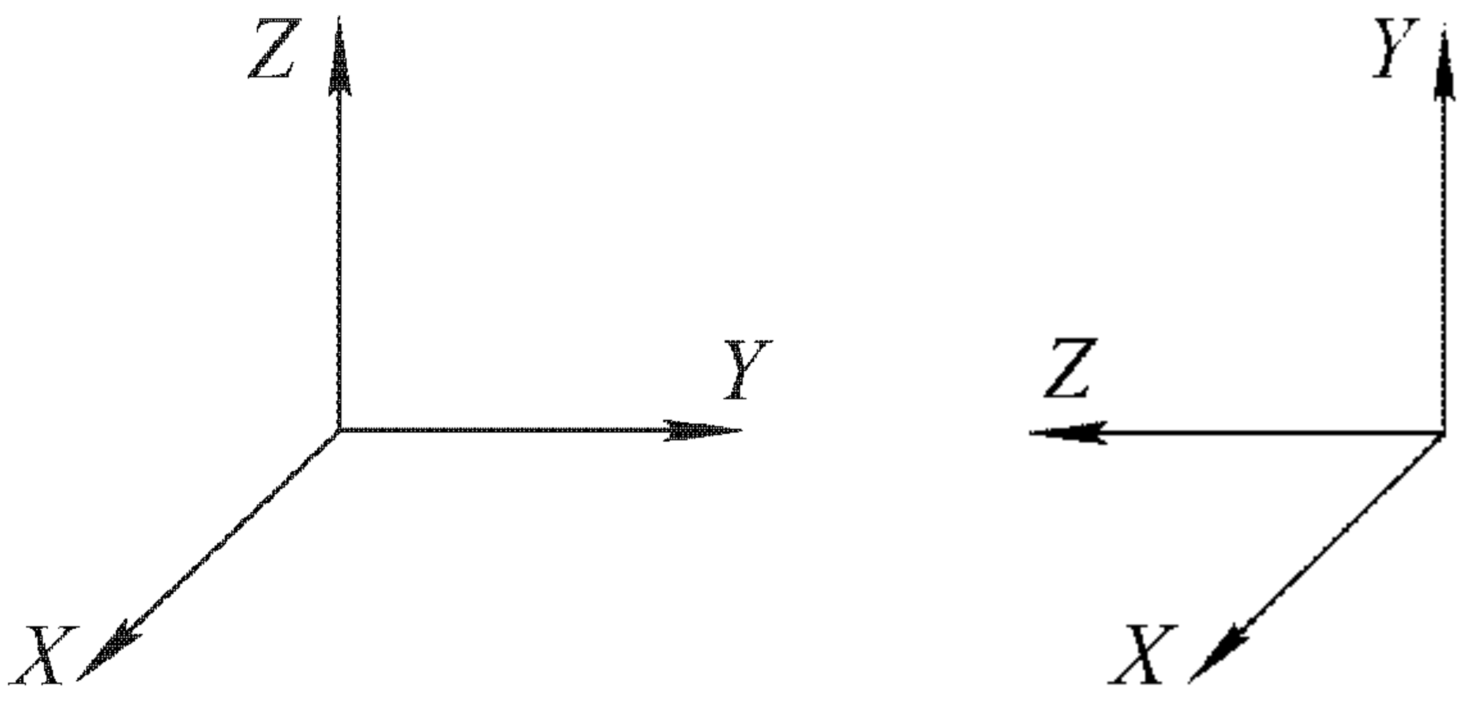


Рис. 320

Что касается коэффициента γ_{xxx}, то он равен нулю. Действительно, в силу симметрии $\gamma_{xxx} = \gamma_{yyy} = \gamma_{zzz}$. Повернем исходную систему координат вокруг оси Z на угол 90°, чтобы ось Y приняла отрицательное направление прежней оси X. Тогда в (96.11), оставляя коэффициенты неизменными, следует сделать замену $x \to -y$, $D_x \to -D_y$, $E_x \to -E_y$, что дает

$$-D_y = -\varepsilon E_y + \gamma_{xxx}\frac{\partial E_y}{\partial y} + \ldots$$

С другой стороны, на основании определения тензора γ_{jlm} в повернутой системе можно сразу написать

$$D_y = \varepsilon E_y + \gamma_{yyy}\frac{\partial E_y}{\partial y} + \ldots$$

Отсюда, ввиду равенства $\gamma_{xxx} = \gamma_{yyy}$, получаем $\gamma_{xxx} = 0$.

Таким образом, все компоненты тензора γ_{jlm} обращаются в нуль, если какие-либо два из индексов j, l, m одинаковы, независимо от значения третьего индекса. Отличны от нуля только компоненты, у которых все три индекса различны. При этом при перестановке любых соседних индексов составляющая тензора γ_{jlm} меняет знак. Следовательно, можно написать

$$\gamma_{xyz} = -\gamma_{yxz} = \gamma_{yzx} = -\gamma_{zyx} = \gamma_{zxy} = -\gamma_{xzy} = -g. \tag{96.12}$$

В результате (96.11) перейдет в

$$D_x = \varepsilon E_x + g\left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right),$$

или в векторной форме

$$\mathbf{D} = \varepsilon\mathbf{E} + g\,\mathrm{rot}\,\mathbf{E}. \tag{96.13}$$

Если каждая точка среды является *центром симметрии*, то при отражении в этом центре среда переходит сама в себя, а потому тензор γ_{jlm} при таком отражении должен оставаться неизменным. Но при этом правая система координат переходит в левую, а знаки координат x, y, z и компонент полярных векторов **E** и **D** меняются на противоположные, так что (96.10) переходит в

$$-D_j = -\varepsilon E_j + \gamma_{jlm}\frac{\partial E_l}{\partial x_m}.$$

Следовательно, $\gamma_{jlm} = 0$, т.е. вращение плоскости поляризации невозможно. Для возможности вращения необходимо, чтобы молекулы жидкости или кристаллов кубической системы *не имели центров симметрии.*

Заметим еще, что в случае изотропной естественно-активной среды величина g есть *псевдоскаляр*, а не истинный скаляр (см. т. I, § 7). При переходе от правой системы координат к левой или наоборот знак этой величины меняется на противоположный. Это непосредственно видно из соотношения (96.12), которое показывает, что γ_{jlm} есть *полностью антисимметричный псевдотензор.*

5. По аналогии с формулой (96.13) можно написать

$$\mathbf{B} = \mathbf{H} + g'\mathrm{rot}\,\mathbf{H}, \tag{96.14}$$

где g' — новый *псевдоскаляр*. Магнитную проницаемость μ мы при этом приняли равной единице. Введение добавочного члена $g'\mathrm{rot}\,\mathbf{H}$ необходимо для выполнения закона сохранения энергии. Действительно, используя уравнения Максвелла, приведем соотношения (96.13) и (96.14) к виду

$$\mathbf{D} = \varepsilon\mathbf{E} - \frac{g}{c}\frac{\partial\mathbf{B}}{\partial t}, \quad \mathbf{B} = \mathbf{H} + \frac{g'}{c}\frac{\partial\mathbf{D}}{\partial t}.$$

Эти уравнения можно упростить. Для этого подставим второе выражение в первое, а первое во второе и отбросим при этом произведение gg', как величину *более высокого порядка малости.* Тогда получим

$$\mathbf{D} = \varepsilon\mathbf{E} - \frac{g}{c}\frac{\partial\mathbf{B}}{\partial t}, \quad \mathbf{B} = \mathbf{H} + \frac{\varepsilon g'}{c}\frac{\partial\mathbf{E}}{\partial t}. \tag{96.15}$$

Теперь воспользуемся результатом электродинамики, согласно которому величина $\mathbf{E}\dot{\mathbf{D}} + \mathbf{H}\dot{\mathbf{B}}$ равна производной по времени от (умноженной на 8π) плотности электромагнитной энергии (см. т. III, § 84). Используя (96.15), преобразуем это выражение к виду

$$\mathbf{E}\dot{\mathbf{D}} + \mathbf{H}\dot{\mathbf{B}} = (\varepsilon\mathbf{E}\dot{\mathbf{E}} + \mathbf{H}\dot{\mathbf{H}}) + \frac{1}{c}(\varepsilon g'\mathbf{E}\ddot{\mathbf{E}} - g\mathbf{H}\ddot{\mathbf{H}}).$$

Первый член справа есть производная от $^1/_2(\varepsilon\mathbf{E}^2 + \mathbf{H}^2)$. Следовательно, и второй член должен быть производной по времени от некоторой функции. Это будет действительно так, если выполняется соотношение $\varepsilon g' = g$, так как тогда

$$\varepsilon g'\mathbf{E}\ddot{\mathbf{E}} - g\mathbf{H}\ddot{\mathbf{H}} = g\frac{\partial}{\partial t}(\mathbf{E}\dot{\mathbf{E}} - \mathbf{H}\dot{\mathbf{H}}).$$

Тем самым доказана необходимость введения второго члена в формуле (96.14), а формулы (96.15) приводятся к окончательному виду

$$\mathbf{D} = \varepsilon\mathbf{E} - \frac{g}{c}\frac{\partial\mathbf{H}}{\partial t}, \quad \mathbf{B} = \mathbf{H} + \frac{g}{c}\frac{\partial\mathbf{E}}{\partial t}. \tag{96.16}$$

6. Теперь мы располагаем полной системой уравнений для монохроматических волн в однородной естественно-активной среде. Из уравнений $\operatorname{div}\mathbf{D} = 0$ и $\operatorname{div}\mathbf{B} = 0$ следует, что плоские волны в такой среде *поперечны относительно векторов* $\mathbf{D}$ *и* $\mathbf{B}$. *Они поперечны также относительно векторов* $\mathbf{E}$ *и* $\mathbf{H}$, так как из уравнений (96.13) и (96.14) следует, что $\operatorname{div}\mathbf{E} = \operatorname{div}\mathbf{H} = 0$. Подставив далее выражения (96.16) в уравнения Максвелла

$$\operatorname{rot}\mathbf{H} = \frac{1}{c}\frac{\partial\mathbf{D}}{\partial t}, \quad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t},$$

получим

$$\begin{aligned} \operatorname{rot}\mathbf{H} &= \frac{\varepsilon}{c}\frac{\partial\mathbf{E}}{\partial t} - \frac{g}{c^2}\frac{\partial^2\mathbf{H}}{\partial t^2}, \\ \operatorname{rot}\mathbf{E} &= -\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t} - \frac{g}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2}. \end{aligned} \tag{96.17}$$

Допустим, что волна плоская и распространяется в положительном направлении оси Z. Тогда отличными от нуля будут только компоненты E_x и E_y, H_x и H_y, причем эти величины зависят только от одной координаты z. С учетом этого запишем уравнения (96.17) в координатной форме

$$\begin{aligned} \frac{\partial E_x}{\partial z} &= -\frac{1}{c}\frac{\partial H_y}{\partial t} - \frac{g}{c^2}\frac{\partial^2 E_y}{\partial t^2}, \\ \frac{\partial E_y}{\partial z} &= \frac{1}{c}\frac{\partial H_x}{\partial t} + \frac{g}{c^2}\frac{\partial^2 E_x}{\partial t^2} \end{aligned} \tag{96.18}$$

и аналогично для производных $\partial H_x/\partial z$ и $\partial H_y/\partial z$. Здесь все величины вещественные. Для упрощения рассуждений удобно ввести комплексные комбинации

$$\begin{aligned} E_+ &= E_x + iE_y, \quad & E_- &= E_x - iE_y, \\ H_+ &= H_x + iH_y, \quad & H_- &= H_x - iH_y. \end{aligned} \tag{96.19}$$

Вещественная часть комплексного числа E_+ дает компоненту E_x, а коэффициент при мнимой части — компоненту E_y, и т.д. Однако при исследовании явлений круговой поляризации удобнее оперировать непосредственно с самими комплексными комбинациями, не переходя к вещественной форме. Например, если совершаются гармонические колебания $E_x = A\cos\omega t$, $E_y = A\sin\omega t$, то $E_+ = Ae^{i\omega t}$. Точка, изображающая комплексное число E_+, движется в комплексной плоскости по кругу в направлении от оси X к оси Y, т.е. представляет *волну, поляризованную по левому кругу*. Аналогично, комплексная комбинация E_- описывает *волну, поляризованную по правому кругу*.

Умножив второе уравнение (96.18) на i, почленным сложением и вычитанием этих уравнений найдем выражения для производных $\partial E_+/\partial z$ и $\partial E_-/\partial z$. Аналогично поступаем с магнитным полем. В результате получим

$$\begin{aligned} \frac{\partial E_+}{\partial z} &= \frac{i}{c}\frac{\partial H_+}{\partial t} + \frac{ig}{c^2}\frac{\partial^2 E_+}{\partial t^2}, \\ \frac{\partial H_+}{\partial z} &= -\frac{i\varepsilon}{c}\frac{\partial E_+}{\partial t} + \frac{ig}{c^2}\frac{\partial^2 H_+}{\partial t^2}. \end{aligned} \tag{96.20}$$

$$\begin{aligned} \frac{\partial E_-}{\partial z} &= -\frac{i}{c}\frac{\partial H_-}{\partial t} - \frac{ig}{c^2}\frac{\partial^2 E_-}{\partial t^2}, \\ \frac{\partial H_-}{\partial z} &= \frac{i\varepsilon}{c}\frac{\partial E_-}{\partial t} - \frac{ig}{c^2}\frac{\partial^2 H_-}{\partial t^2}. \end{aligned} \tag{96.21}$$

Уравнения разделились на две группы *независимых уравнений*. В одну группу входят комбинации E_+ и H_+, описывающие волны с левой круговой поляризацией, в другую — комбинации E_- и H_-, которым соответствует правая круговая поляризация. Ввиду однотипности обеих групп достаточно исследовать одну из них, например группу (96.20). Исключим из нее стандартным способом величину H_+. Отбрасывая члены, содержащие g^2, получим

$$\frac{\partial^2 E_+}{\partial z^2} = \frac{\varepsilon}{c^2}\frac{\partial^2 E_+}{\partial t^2} + \frac{2ig}{c^2}\frac{\partial^3 E_+}{\partial t^2\,\partial z}$$

и такое же уравнение для H_+. В монохроматической плоской волне $E_+ = C\exp\left[i(\omega t - kz)\right]$ дифференцирование по z эквивалентно умножению на $-ik$. Поэтому

$$\frac{\partial^2 E_+}{\partial z^2} = \frac{\varepsilon}{c^2}\frac{\partial^2 E_+}{\partial t^2} + \frac{2gk}{c^2}\frac{\partial^2 E_+}{\partial t^2}.$$

Здесь в последнем члене вместо k можно подставить волновое число в нулевом приближении, т.е. $k = \omega\sqrt{\varepsilon}/c$. Тогда получится волновое уравнение

$$\frac{\partial^2 E_+}{\partial z^2} = \frac{1}{v_+^2}\frac{\partial^2 E_+}{\partial t^2}, \tag{96.22}$$

в котором

$$v_+ = \frac{c}{\sqrt{\varepsilon}}\frac{1}{\left[1 + 2g\omega/c\sqrt{\varepsilon}\right]^{1/2}} = \frac{c}{\sqrt{\varepsilon}} - \frac{g\omega}{\varepsilon},$$

или

$$v_+ = \frac{c}{n} - \frac{g\omega}{n^2}, \tag{96.23}$$

где n — показатель преломления среды. Аналогично, для волны с левой круговой поляризацией

$$v_- = \frac{c}{n} + \frac{g\omega}{n^2}. \tag{96.24}$$

Скорости v_+ и v_- различны. Поэтому должно существовать *круговое двойное лучепреломление*, а следовательно, и *вращение плоскости поляризации.*

Убедимся еще, что в каждой поляризованной по кругу плоской волне электрический и магнитный векторы взаимно перпендикулярны. Для этого в уравнениях (96.20) достаточно произвести замену $\partial/\partial t \to i\omega$, $\partial/\partial z \to -ik$. Тогда первое уравнение перейдет в

$$\frac{\omega}{c}H_+ = i\left(k - \frac{g\omega^2}{c}\right)E_+.$$

Отсюда видно, что отношение E_+/H_+ *чисто мнимое*, а это эквивалентно утверждению, что векторы $\mathbf{E}$ и $\mathbf{H}$ в рассматриваемой волне взаимно перпендикулярны.

7. В средах, обладающих *центром симметрии*, разложение (96.9) тензора $\varepsilon_{jl}(\omega, \mathbf{k})$ не может содержать линейных членов по $\mathbf{k}$. В таких случаях пространственная дисперсия может быть обусловлена квадратичными членами. С ней связана слабая зависимость поглощения кубических кристаллов закиси меди Cu_2O от поляризации света (Гросс и Каплянский, 1960 г.), а также слабая анизотропия показателя преломления кубических кристаллов кремния (Пастернак и Ведам, 1971 г.). Эффекты очень малы, так как они определяются квадратом $(a/\lambda)^2$, т.е. величиной порядка 10^{-6}–10^{-5} (a — размер молекулы или постоянная кристаллической решетки). Существуют и другие эффекты пространственной дисперсии, в обсуждение которых мы входить не будем.

§ 97. О тепловых флуктуациях

В этом вспомогательном параграфе сообщаются краткие сведения о тепловых флуктуациях, необходимые для изложения вопросов молекулярного рассеяния света.

1. Пусть f — любая флуктуирующая величина. *Флуктуацией* этой величины называется отклонение $\Delta f = f - \overline{f}$ мгновенного значения этой величины от ее среднего значения. Обычно пользуются *средним квадратом флуктуации*, т.е. величиной $\overline{(\Delta f)^2}$. Квадратный корень из этой величины $\sqrt{\overline{(\Delta f)^2}}$ называется *среднеквадратичной флуктуацией*, а ее отношение к среднему значению $\overline{f}$, т.е. $\sqrt{\overline{(\Delta f)^2}}/\overline{f}$, — *среднеквадратичной относительной флуктуацией.*

Усредняя выражение $(\Delta f)^2 = (f - \overline{f})^2 = f^2 - 2f\overline{f} + (\overline{f})^2$, получим

$$\overline{(\Delta f)^2} = \overline{f^2} + 2\overline{f\overline{f}} - (\overline{f})^2.$$

Но $\overline{f}$ — величина постоянная, а потому $\overline{f\overline{f}} = \overline{f}\,\overline{f} = (\overline{f})^2$. Следовательно,

$$\overline{(\Delta f)^2} = \overline{f^2} - (\overline{f})^2. \tag{97.1}$$

Усредним теперь произведение двух флуктуирующих величин:

$$fg = (\overline{f} + \Delta f)(\overline{g} + \Delta g) = \overline{f}\,\overline{g} + \overline{g}\,\Delta f + \overline{f}\,\Delta g + \Delta f\,\Delta g.$$

Так как $\overline{f}$ и $\overline{g}$ — величины постоянные, а $\overline{\Delta f} = \overline{\Delta g} = 0$, то

$$\overline{fg} = \overline{f}\,\overline{g} + \overline{\Delta f\,\Delta g}. \tag{97.2}$$

Формула (97.1) содержится здесь как частный случай, если взять $f = g$.

Величины f и g называются *статистически независимыми*, если $\overline{\Delta f\,\Delta g} = 0$. Для таких величин

$$\overline{fg} = \overline{f}\,\overline{g}. \tag{97.3}$$

2. Рассмотрим теперь любую физическую систему, состоящую из N независимых одинаковых частей. Примером такой системы может служить идеальный газ, а составных частей — отдельные молекулы. Пусть f_i — произвольная *аддитивная величина*, характеризующая i-ю подсистему, например в приведенном примере — кинетическая энергия i-й молекулы. Тогда в силу предполагаемой аддитивности соответствующая величина для всей системы будет $F = \sum f_i$. Выразим средний квадрат флуктуации величины F через аналогичный квадрат для величины f_i. Очевидно, $\overline{F} = \sum \overline{f}_i = N\overline{f}$, где опущен индекс i, так как предполагается, что все составные части системы тождественны. Далее,

$$F^2 = \left(\sum f_i\right)^2 = \sum f_i^2 + \sum_{i \neq j}\sum f_i f_j.$$

А так как эти части независимы, то $\overline{f_i f_j} = \overline{f}_i \overline{f}_j = (\overline{f})^2$. Следовательно,

$$\overline{F^2} = N\overline{f^2} + N(N-1)(\overline{f})^2.$$

Подставляя эти значения в формулу (97.1), получим

$$\overline{(\Delta F)^2} = \overline{F^2} - (\overline{F})^2 = N(\overline{f^2} - \overline{f}^2). \tag{97.4}$$

Отсюда на основании (97.1)

$$\frac{\sqrt{\overline{(\Delta F)^2}}}{\overline{F}} = \frac{\sqrt{N}\sqrt{\overline{(\Delta f)^2}}}{N\overline{f}} = \frac{1}{\sqrt{N}}\frac{\sqrt{\overline{(\Delta f)^2}}}{\overline{f}}. \tag{97.5}$$

Таким образом, относительная флуктуация величины F обратно пропорциональна квадратному корню из N. При больших N относительные флуктуации ничтожны. Этот вывод верен и для неаддитивных величин. С ним связана *достоверность термодинамических результатов для макроскопических систем.*

3. Применим формулу (97.4) к вычислению флуктуаций числа молекул в фиксированном объеме V идеального газа. Пусть в большом закрытом сосуде объема V в отсутствие силовых полей находятся N молекул идеального газа. Разделим объем V на $z = V/v$ одинаковых объемчиков v_i величиной v каждый. Если n_i — число молекул в объемчике v_i, а N — в большом объеме V, то $N = \sum n_i$. Среднее число молекул $\overline{n}_i$ в объемчике v_i будет $\overline{n}_i = \overline{n} = Nv/V$, т.е. одно и то же во всех объемчиках. Представим его в виде $\overline{n} = Np$, где $p = v/V$ — вероятность нахождения молекулы в объемчике v. Возьмем теперь в качестве f_i следующие величины: $f_i = 1$, если i-я молекула находится внутри объемчика v, и $f_i = 0$, если она находится в оставшемся объеме $V - v$. Тогда число молекул n в объемчике v можно представить в виде $n = \sum f_i$, предполагая, что суммирование ведется по всем N молекулам объема V. Ясно, что $f_i = f_i^2 = f_i^3 = \ldots$, а потому $\overline{f}_i = \overline{f_i^2} = \overline{f_i^3} = \ldots = p$. Следовательно, по формуле (97.1)

$$\overline{\Delta f_i^2} = \overline{f_i^2} - (\overline{f}_i)^2 = p - p^2 = p(1-p).$$

А так как в случае идеального газа величины f_1, f_2, f_3, … статистически независимы, то по формуле (97.4)

$$\overline{\Delta n^2} = Np(1-p) = (1-p)\overline{n}. \tag{97.6}$$

Если объем V бесконечно велик, то $p \to 0$, а следовательно,

$$\overline{\Delta n^2} = \overline{n}. \tag{97.7}$$

Отсюда получаем для относительной флуктуации плотности газа

$$\sqrt{\overline{\Delta\rho^2}}/\overline{\rho} = 1/\sqrt{\overline{n}}. \tag{97.8}$$

4. Более общий метод вычисления флуктуаций плотности, применимый также к жидкостям и твердым телам, основан на *теореме о равномерном* *распределении кинетической энергии по степеням свободы.* Рассмотрим малую часть жидкости или газа, окруженную такой же жидкой или газообразной средой, температура которой T поддерживается постоянной (термостатом). С целью упрощения и наглядности вычислений предположим, что эта малая часть жидкости или газа заключена в цилиндр с поршнем. Стенки цилиндра идеально проводят тепло, а поршень может ходить в нем без трения. Тогда наличие стенок цилиндра и поршня не будет препятствовать обмену энергией и выравниванию давлений между веществом в цилиндре и термостатом. Благодаря тепловому движению поршень будет совершать броуновское движение. К нему мы и применим теорему о равномерном распределении кинетической энергии по степеням свободы.

Поршень можно рассматривать как гармонический осциллятор, совершающий тепловые колебания. Среднее значение его потенциальной энергии при смещении на x из положения равновесия $x = 0$ равно ${}^1\!/_2\varkappa\overline{x^2} = {}^1\!/_2kT$, где $\varkappa$ — коэффициент упругости, соответствующий такому смещению. Если S — площадь поршня, а ΔV — изменение объема системы, то $\Delta V = Sx$. Таким образом, $\overline{(\Delta V)^2} = S^2\overline{x^2} = S^2kT/\varkappa$. Сила, возвращающая поршень в положение равновесия, будет $F = S\dfrac{\partial P}{\partial x}x$, где P — давление газа или жидкости. Поэтому $\varkappa = -S\,\partial P/\partial x = -S^2\partial P/\partial V$. В результате получим

$$(\overline{\Delta V^2})_T = -\frac{kT}{(\partial P/\partial V)_T} = -kT(\partial V/\partial P)_T. \qquad (97.9)$$

Значок T указывает, что в выводе предполагалось *постоянство температуры* окружающей среды (термостата). Если бы вещество внутри объема V было адиабатически изолировано, то T следовало бы заменить на S (постоянство энтропии), т.е.

$$(\overline{\Delta V^2})_S = -kT(\partial V/\partial P)_S. \qquad (97.10)$$

Формулы (97.9) и (97.10) выражают флуктуации объема *одной и той же массы вещества*, находящейся в термодинамическом равновесии с окружающей средой. Для идеального газа при постоянстве температуры $PV = \text{const}$, так что $(\partial V/\partial P)_T = -V/P$. А так как $PV = nkT$, где n — число молекул в объеме V, то из формулы (97.9) получаем $\overline{(\Delta V)^2} = V^2/n$. Пусть теперь объем V фиксирован, а число молекул в нем n меняется из-за тепловых флуктуаций. Если бы объем V увеличился на ΔV, то число молекул в прежнем объеме V уменьшилось бы на $\Delta n = n\,\Delta V/V$. Отсюда

$$\overline{(\Delta n)^2} = (n/V)^2\overline{(\Delta V)^2} = n = \overline{n},$$

что совпадает с прежним результатом (97.7). В окончательном результате различием между n и $\overline{n}$ можно пренебречь, что мы и сделали.

5. Перейдем теперь к вычислению *флуктуаций энергии.* С целью лучшего уяснения метода начнем с вычисления флуктуаций кинетической энергии $\mathcal{E}$ молекулы одноатомного идеального газа. Согласно максвелловскому закону распределения скоростей,

$$\overline{\mathcal{E}} = \frac{1}{Z}\int \mathcal{E}e^{-\alpha\mathcal{E}}\,d\Gamma, \qquad (97.11)$$

где $\alpha = 1/(kT)$, $d\Gamma$ — элемент объема пространства скоростей, а Z определяется условием нормировки:

$$Z = \int e^{-\alpha\mathcal{E}}\,d\Gamma. \qquad (97.12)$$

Дифференцируя это соотношение по параметру α, получим

$$\frac{dZ}{d\alpha} = -\int \mathcal{E} e^{-\alpha\mathcal{E}}\, d\Gamma,$$

и формула (97.11) перейдет в

$$\overline{\mathcal{E}} = -\frac{1}{Z}\frac{dZ}{d\alpha}. \qquad (97.13)$$

Отсюда

$$\frac{d\overline{\mathcal{E}}}{d\alpha} = -\frac{1}{Z}\frac{d^2Z}{d\alpha^2} + \frac{1}{Z^2}\left(\frac{dZ}{d\alpha}\right)^2 = -\frac{1}{Z}\frac{d^2Z}{d\alpha^2} + \overline{\mathcal{E}}.$$

Аналогично,

$$\overline{\mathcal{E}^2} = \frac{1}{Z}\int \mathcal{E}^2 e^{-\alpha\mathcal{E}}\, d\Gamma = \frac{1}{Z}\frac{d^2Z}{d\alpha^2}. \qquad (97.14)$$

Сравнение этой формулы с предыдущей дает

$$\overline{(\Delta\mathcal{E})^2} = \overline{\mathcal{E}^2} - (\overline{\mathcal{E}})^2 = -\frac{d\overline{\mathcal{E}}}{d\alpha},$$

или после подстановки значения параметра $\alpha = 1/(kT)$

$$\overline{(\Delta\mathcal{E})^2} = kT^2\frac{d\overline{\mathcal{E}}}{dT}. \qquad (97.15)$$

Так как для идеального одноатомного газа $\overline{\mathcal{E}} = {}^3\!/_2 kT$, то

$$\overline{(\Delta\mathcal{E})^2} = \frac{3}{2}(kT)^2. \qquad (97.16)$$

6. Приведенный метод вычисления $\overline{(\Delta\mathcal{E})^2}$ может быть распространен практически без изменений на более общие и важные случаи. Выделим, например, малую часть (подсистему) изотропной среды (жидкости или газа), находящуюся в статистическом равновесии со всей средой, температура T которой поддерживается постоянной. По отношению к выделенной подсистеме окружающая среда играет роль *термостата*. Из-за обмена энергией между термостатом и подсистемой энергия последней будет непрерывно флуктуировать. Беспорядочные изменения энергии подсистемы подчиняются статистическому закону, вполне аналогичному максвелловскому закону распределения кинетической энергии между молекулами. Поэтому среднее значение энергии подсистемы будет выражаться прежней формулой (97.13), где α имеет прежнее значение, а интегрирование в выражении (97.12) производится по многомерному пространству координат и импульсов подсистемы. В этом единственное отличие рассматриваемого случая от предыдущего. Но оно совсем не отражается на последующих выкладках. Поэтому окончательный результат (97.15) применим к рассматриваемой подсистеме без всяких изменений.

Результат (97.15) остается верным и в квантовой статистике. Чтобы убедиться в этом, достаточно заменить интеграл (97.12) суммой

$$Z = \sum g_i \exp(-\alpha\mathcal{E}_i),$$

где $\mathcal{E}_i$ — энергии квантовых уровней подсистемы, а g_i — кратности этих уровней. Суммирование производится по всем квантовым состояниям подсистемы (см. т. II, § 85).

В случае макроскопической подсистемы, объем которой поддерживается постоянным, $\mathcal{E}$ имеет смысл *внутренней энергии подсистемы*, а $d\mathcal{E}/dT$ — ее *теплоемкости* C_V при постоянном объеме. Тогда из формулы (97.15) получается

$$(\overline{\Delta\mathcal{E}^2})_V = kT^2C_V. \tag{97.17}$$

Значок V слева указывает на то, что величина $(\overline{\Delta\mathcal{E}^2})_V$ есть средний квадрат флуктуации энергии подсистемы при сохранении ее объема V постоянным.

7. Рассмотрим теперь *флуктуации энтальпии* I подсистемы. Для этого воспользуемся следующим искусственным приемом. Предположим, что подсистема заключена в оболочку с идеально проводящими подвижными стенками, так что объем подсистемы не сохраняется постоянным. Пусть оболочка снаружи подвергается действию постоянных внешних сил, поддерживающих внешнее давление P постоянным. Эти силы увеличивают потенциальную энергию подсистемы на величину PV. Если под $\mathcal{E}$ понимать ту же энергию, что и в предыдущем выводе, то с учетом дополнительной потенциальной энергии PV среднее значение полной энергии подсистемы будет $\overline{\mathcal{E}}+P\overline{V}$. Но это есть энтальпия подсистемы I. Все предыдущие рассуждения можно повторить без изменений, заменив $\mathcal{E}$ на $\mathcal{E}+PV$. В результате вместо формулы (97.15) получится

$$(\overline{\Delta I^2})_P = kT^2\frac{dI}{dT}. \tag{97.18}$$

Но при $P = \text{const}$ производная dI/dT есть теплоемкость C_P подсистемы при постоянном давлении, а потому

$$(\overline{\Delta I^2})_P = kT^2C_P. \tag{97.19}$$

8. Распространим теперь термодинамический метод вычисления флуктуаций, изложенный выше, на *любые величины*, характеризующие макроскопические свойства подсистем. Ограничимся при этом изотропными телами. Для них любая термодинамическая величина в состоянии термодинамического равновесия есть функция двух других термодинамических величин, которые могут быть приняты за *независимые переменные*. Термодинамические величины макроскопических подсистем хотя и испытывают флуктуации, но в случае малости таких подсистем их мгновенные состояния *практически равновесны*. Они также определяются двумя независимыми переменными. Поэтому задача сводится к вычислению тепловых флуктуаций таких двух независимых переменных. В окончательном результате, определяющем значение среднего квадрата той или иной флуктуации, необходимо указывать, какая из двух величин, выбранных для характеристики состояния подсистемы, *поддерживается постоянной*. Иначе самый результат будет неопределенным, а потому и бессмысленным.

Покажем на примерах, как применяется изложенный метод к вычислению флуктуаций различных физических величин.

Начнем с флуктуаций температуры, предполагая, что рассматриваемая макроскопическая подсистема находится в тепловом контакте с термостатом. Считая температуру подсистемы функцией независимых переменных V и $\mathcal{E}$, напишем

$$\Delta T = \left(\frac{\partial T}{\partial V}\right)_{\mathcal{E}}\Delta V + \left(\frac{\partial T}{\partial \mathcal{E}}\right)_V\Delta\mathcal{E}.$$

В силу независимости V и $\mathcal{E}$ имеем: $\overline{\Delta V\cdot\Delta\mathcal{E}} = 0$, а потому

$$\overline{(\Delta T)^2} = \left(\frac{\partial T}{\partial V}\right)_{\mathcal{E}}^2\overline{(\Delta V)^2} + \left(\frac{\partial T}{\partial \mathcal{E}}\right)_V^2\overline{(\Delta\mathcal{E})^2}.$$

При постоянном объеме производная $(\partial\mathcal{E}/\partial T)_V$ есть теплоемкость подсистемы C_V. Тогда из формулы (97.17) находим

$$(\overline{\Delta T^2})_V = \frac{1}{C_V^2}(\overline{\Delta\mathcal{E}^2})_V = \frac{kT^2}{C_V}. \tag{97.20}$$

Вычислим теперь флуктуации энтропии S подсистемы. В качестве независимых переменных выберем V и $\mathcal{E}$. Рассуждая как в предыдущем примере, напишем

$$\overline{\Delta V{\cdot}\Delta\mathcal{E}} = 0, \quad (\overline{\Delta S^2})_V = \left(\frac{\partial S}{\partial\mathcal{E}}\right)_V^2 (\overline{\Delta\mathcal{E}^2})_V.$$

Так как $\partial S/\partial\mathcal{E} = 1/T$, то с учетом (97.17)

$$(\overline{\Delta S^2})_V = kC_V. \tag{97.21}$$

Если бы за независимые переменные были приняты P и I, то получилось бы

$$(\overline{\Delta S^2})_P = kC_P. \tag{97.22}$$

Рассмотрим теперь флуктуации давления P. Примем за независимые переменные V и T. Тогда

$$\overline{\Delta V\cdot\Delta T} = 0, \quad (\overline{\Delta P^2})_T = \left(\frac{\partial P}{\partial V}\right)_T^2 (\overline{\Delta V^2})_T,$$

или на основании формулы (97.9)

$$(\overline{\Delta P^2})_T = -kT\left(\frac{\partial P}{\partial V}\right)_T. \tag{97.23}$$

Приняв за независимые переменные P и S, мы получили бы

$$(\overline{\Delta P^2})_S = -kT\left(\frac{\partial P}{\partial V}\right)_S = -\gamma kT\left(\frac{\partial P}{\partial V}\right)_T, \tag{97.24}$$

где $\gamma = C_P/C_V$. При этом было использовано термодинамическое соотношение $\left(\frac{\partial P}{\partial V}\right)_S \Big/ \left(\frac{\partial P}{\partial V}\right)_T = \gamma$ (см.т. II, § 47).

Вычислим, наконец, *флуктуации плотности* вещества ρ в объеме V. Задача сводится просто к преобразованию формулы (97.9). Прежде всего заметим, что величина $V\,\partial P/\partial V$ не зависит от величины объема V. Поэтому в таком выражении V можно заменить удельным объемом вещества v. Тогда

$$(\overline{\Delta V^2})_T = -\frac{VkT}{v(\partial P/\partial v)_T}.$$

Так как масса вещества $V\rho$ в объеме V остается постоянной, то $V\,\Delta\rho + \rho\,\Delta V = 0$. Следовательно, $(\Delta\rho)^2 = \left(\frac{\rho}{V}\right)^2(\Delta V)^2$, а потому

$$V(\overline{\Delta\rho^2})_T = -\frac{\rho^2kT}{v(\partial P/\partial v)_T}. \tag{97.25}$$

Чем меньше объем V, тем больше относительные флуктуации плотности в нем.

§ 98. Рассеяние света

1. В прозрачной однородной среде бегущая плоская волна распространяется только в прямом направлении, не испытывая рассеяния в стороны. (Мы отвлекаемся от дифракции, предполагая, что ширина фронта волны достаточно велика, а следовательно, угол дифракционной расходимости мал.) Допустим теперь, что оптическая однородность среды нарушена, например множеством мельчайших частиц постороннего вещества, беспорядочно распределенных по объему среды. Примерами могут служить пыльный воздух, туман, дым, эмульсии и суспензии с взвешенными в них посторонними частицами. Тогда показатель преломления будет меняться в пространстве весьма нерегулярно, но среднее значение его во всяком малом объеме, содержащем еще очень много макроскопических неоднородностей, будет оставаться одним и тем же во всей среде. Такую среду называют *оптически мутной.* В оптически мутных средах свет распространяется не только в прямом направлении, но и *рассеивается в стороны.* Рассеяние света в мутных средах на частицах постороннего вещества экспериментально впервые исследовал Тиндаль (1820–1893) в 1869 г. Поэтому это явление получило название *тиндалевского рассеяния* или *эффекта Тиндаля.* Его теория была дана Рэлеем.

2. В неоднородной неподвижной изотропной среде распространение света описывается уравнениями Максвелла

$$\begin{aligned} \operatorname{rot}\mathbf{H} &= \frac{\varepsilon}{c}\frac{\partial\mathbf{E}}{\partial t}, \qquad \operatorname{div}(\varepsilon\mathbf{E}) = 0, \\ \operatorname{rot}\mathbf{E} &= -\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t}, \qquad \operatorname{div}\mathbf{H} = 0, \end{aligned} \tag{98.1}$$

где диэлектрическая проницаемость ε является функцией координат. Выделим из нее постоянную часть ε_0, полагая $\varepsilon = \varepsilon_0 + \delta\varepsilon$. В проблеме рассеяния света интерес представляет случай, когда $\delta\varepsilon$ мало по сравнению с ε_0, но пока мы не будем вводить этого ограничения. Более того, постоянное слагаемое ε_0 в принципе можно было бы выбрать произвольно. От этого, если вычисления производить точно, окончательный результат зависеть не может. Однако удобно и естественно понимать под ε_0 диэлектрическую проницаемость среды, из которой удалены частицы постороннего вещества.

Представим электромагнитное поле в виде $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}'$, $\mathbf{H} = \mathbf{H} + \mathbf{H}'$, где $\mathbf{E}_0$, $\mathbf{H}_0$ удовлетворяют уравнениям Максвелла в однородной среде

$$\begin{aligned} \operatorname{rot}\mathbf{H}_0 &= \frac{\varepsilon_0}{c}\frac{\partial\mathbf{E}_0}{\partial t}, \qquad \operatorname{div}(\varepsilon_0\mathbf{E}_0) = 0, \\ \operatorname{rot}\mathbf{E}_0 &= -\frac{1}{c}\frac{\partial\mathbf{H}_0}{\partial t}, \qquad \operatorname{div}\mathbf{H}_0 = 0. \end{aligned}$$

В задаче о рассеянии света это есть *падающая волна*, которая распространялась бы в среде, если бы в ней не было оптических неоднородностей, а $\mathbf{E}'$, $\mathbf{H}'$ — поле *рассеянного света*. Вычитая предыдущие уравнения из (98.1), получим

$$\begin{aligned} &\operatorname{rot}\mathbf{H}' - \frac{\varepsilon_0}{c}\frac{\partial \mathbf{E}'}{\partial t} = \frac{\delta\varepsilon}{c}\frac{\partial \mathbf{E}}{\partial t}, \qquad \operatorname{div}(\varepsilon_0\mathbf{E}') = -\operatorname{div}(\delta\varepsilon\mathbf{E}),\\ &\operatorname{rot}\mathbf{E}' - \frac{1}{c}\frac{\partial \mathbf{H}'}{\partial t} = 0, \qquad \operatorname{div}\mathbf{H}' = 0. \end{aligned} \tag{98.2}$$

Таким образом, для поля $\mathbf{E}'$, $\mathbf{H}'$ получились такие же уравнения Максвелла, как в однородной среде с диэлектрической проницаемостью ε_0. Только первые два из этих уравнений содержат правые части, которые можно рассматривать как *дополнительные источники электромагнитных волн*. Если ввести обозначение

$$\delta\mathbf{P} = \frac{\delta\varepsilon}{4\pi}\mathbf{E}, \tag{98.3}$$

то эти два уравнения перейдут в

$$\operatorname{rot}\mathbf{H}' - \frac{\varepsilon_0}{c}\frac{\partial \mathbf{E}'}{\partial t} = \frac{4\pi}{c}\frac{\partial}{\partial t}\delta\mathbf{P}, \qquad \operatorname{div}(\varepsilon_0\mathbf{E}') = -4\pi\operatorname{div}(\delta\mathbf{P}). \tag{98.4}$$

Из них видно, что в среде появляется *дополнительная поляризация* $\delta\mathbf{P}$, определяемая выражением (98.3), так что каждый малый элемент объема среды δV получает *дополнительный дипольный момент* $\delta V\cdot\delta\mathbf{P}$. Меняясь во времени, он излучает электромагнитные волны как колеблющийся *диполь Герца*. Это и есть свет, рассеянный элементом объема δV.

3. Допустим теперь, что оптическая неоднородность создается одинаковыми шариками радиуса a, беспорядочно распределенными по объему, занятому средой. Пусть среднее расстояние между шариками велико по сравнению с a, а сами шарики малы по сравнению с длиной волны. Тогда при вычислении электрического поля $\mathbf{E}$ внутри шарика можно считать внешнее поле $\mathbf{E}_0$ световой волны однородным. Как показано в электростатике (см. т. III, § 16), поле $\mathbf{E}$ также однородно и определяется выражением

$$\mathbf{E} = \frac{3}{\varepsilon/\varepsilon_0 + 2}\mathbf{E}_0 = \frac{3\varepsilon_0}{\varepsilon + 2\varepsilon_0}\mathbf{E}_0, \tag{98.5}$$

где ε — диэлектрическая проницаемость шарика, а ε_0 — окружающей среды. Дополнительная поляризация, согласно формуле (98.3), будет отлична от нуля только внутри шариков, где она равна

$$\delta\mathbf{P} = \frac{\varepsilon - \varepsilon_0}{4\pi}\mathbf{E} = \frac{\varepsilon - \varepsilon_0}{4\pi}\frac{3\varepsilon_0}{\varepsilon + 2\varepsilon_0}\mathbf{E}_0,$$

а дополнительный дипольный момент шарика

$$\mathbf{p} = \frac{\varepsilon_0}{\varepsilon + 2\varepsilon_0}(\varepsilon - \varepsilon_0)a^3\mathbf{E}_0. \tag{98.6}$$

Предположим сначала, что падающая волна *поляризована линейно.* Тогда векторы **p** и **E** все время будут параллельны одному и тому же неизменному направлению. Электрическое поле диполя **p** на больших расстояниях r от него (в волновой зоне) определяется выражением

$$E_1 = \frac{\sin\vartheta}{c^2 r}[\ddot{p}]_{t-r/v} = -\frac{\omega^2\sin\vartheta}{c^2 r}[p]_{t-r/v}, \tag{98.7}$$

где $v = c/\sqrt{\varepsilon} = c/n$ — скорость света в рассматриваемой среде, а ϑ — угол между осью диполя **p** и направлением рассеянного излучения (см. т. III, § 141). *Рассеянный свет поляризован линейно*, причем электрический вектор лежит в плоскости, проходящей через ось диполя **p** и направление излучения. *Под интенсивностью света здесь и в дальнейшем будем понимать усредненное по времени численное значение вектора Пойнтинга.* Для интенсивности света, рассеянного одним шариком, электродинамика дает

$$I_1 = \frac{\sin^2\vartheta}{4\pi\varepsilon_0 v^3 r^2}\overline{\ddot{p}^2} = \frac{\omega^4\sin^2\vartheta}{4\pi\varepsilon_0 v^3 r^2}\overline{p^2} \tag{98.8}$$

(см. т. III, § 141). Интенсивность прямой волны равна

$$I_0 = \frac{c}{4\pi}\overline{E_0 H_0} = \frac{v}{4\pi}\varepsilon_0\overline{E_0^2}. \tag{98.9}$$

Воспользовавшись выражением (98.6), получим

$$I_1 = 9\varepsilon_0^2\left(\frac{\varepsilon-\varepsilon_0}{\varepsilon+2\varepsilon_0}\right)^2\frac{\omega^4 a^6\sin^2\vartheta}{c^4 r^2}I_0, \tag{98.10}$$

или

$$I_1 = 9\varepsilon_0^2\left(\frac{\varepsilon-\varepsilon_0}{\varepsilon+2\varepsilon_0}\right)^2\frac{\pi^2 V_1^2}{\lambda^4}\frac{\sin^2\vartheta}{r^2}I_0, \tag{98.11}$$

где λ — длина волны в вакууме, а $V_1 = {}^4\!/_3\pi a^3$ — объем шарика. Энергия $\mathcal{P}_1$, рассеиваемая шариком в единицу времени по всем направлениям, найдется интегрированием величины (98.11) по сфере радиуса r. Взяв в качестве элемента поверхности $2\pi r^2\sin\vartheta\,d\vartheta$, получим

$$\mathcal{P}_1 = 24\pi^3\varepsilon_0^2\left(\frac{\varepsilon-\varepsilon_0}{\varepsilon+2\varepsilon_0}\right)^2\frac{V_1^2}{\lambda^4}I_0. \tag{98.12}$$

Допустим теперь, что падающий свет *естественный.* Направление его распространения примем за ось Z. Пусть рассеянный свет наблюдается в направлении OA под углом θ к оси Z. Угол θ называется *углом рассеяния* (рис. 321). Направим ось X перпендикулярно к OA и OZ. Так как **p** и $\mathbf{E}_0$ коллинеарны, то вектор **p** параллелен плоскости XY. Разложим его по осям X и Y. Интенсивности излучений дипольных моментов $\mathbf{p}_x$ и $\mathbf{p}_y$ найдутся по формуле (98.8), если в ней положить сначала $\vartheta = \pi/2$, а затем $\vartheta = \frac{\pi}{2} - \theta$. Так как падающий свет естественный, то эти излучения *некогерентны*, так что для нахождения I_1 надо сложить их интенсивности. В результате формула (98.8) перейдет в

$$I_1 = \frac{\omega^4}{4\pi\varepsilon_0 v^3 r^2}\left(\overline{p_x^2} + \overline{p_y^2}\cos^2\theta\right) = \frac{\omega^4}{4\pi\varepsilon_0 v^3 r^2}\,\frac{1+\cos^2\theta}{2}\,\overline{p^2},$$

так как в случае естественного света $\overline{p_x^2} = \overline{p_y^2} = {}^1\!/_2\overline{p^2}$. Следовательно, вместо формулы (98.11) получится

$$I_1 = 9\varepsilon_0^2\left(\frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0}\right)^2 \frac{\pi^2 V_1^2}{\lambda^4}\,\frac{1+\cos^2\theta}{2r^2}\,I_0. \tag{98.13}$$

Формула (98.12), очевидно, останется без изменения.

Теперь рассеянный свет будет поляризован частично. *Полная линейная поляризация будет наблюдаться только в тех случаях, когда линия наблюдения OA перпендикулярна к направлению распространения падающего света*, так как в этом случае дипольный момент $\mathbf{p}_y$ излучения не дает.

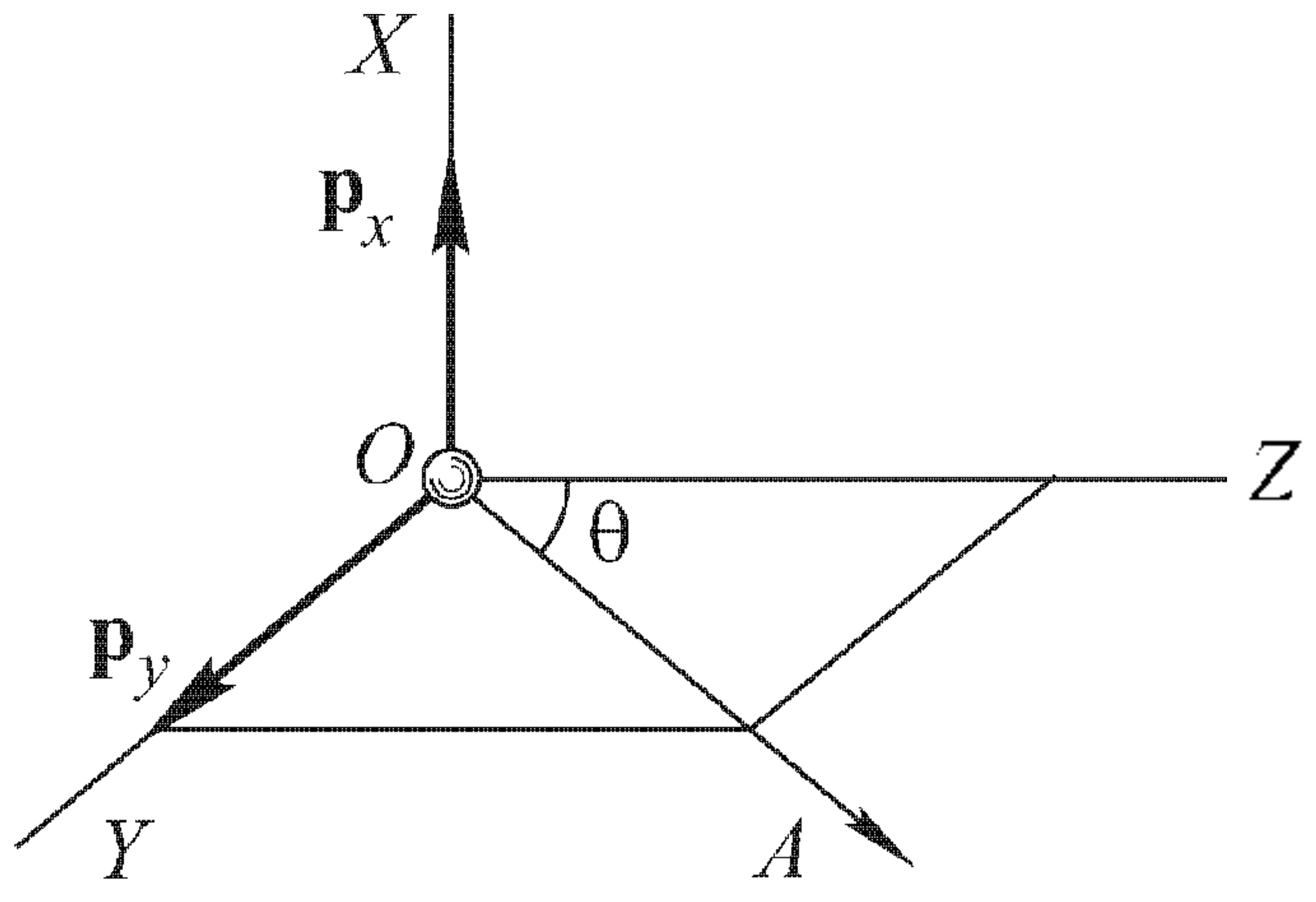


Рис. 321

Найдем теперь интенсивность I света, рассеиваемого объемом V, содержащим очень много шариков. Их среднее число в этом объеме равно NV, где N — среднее число шариков в единице объема. Так как расстояния между шариками велики по сравнению с a и они распределены по объему V беспорядочно, то для нахождения I надо сложить интенсивности, рассеиваемые отдельными шариками. Предположим, что расстояние от объема V до точки наблюдения велико по сравнению с линейными размерами самого объема V. Тогда в формуле (98.13) все расстояния r можно считать одинаковыми и написать

$$I = 9\varepsilon_0^2\left(\frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0}\right)^2 \pi^2\frac{V_1^2}{\lambda^4}\,\frac{1+\cos^2\theta}{2r^2}\,NVI_0. \tag{98.14}$$

4. Рассчитаем убывание интенсивности I_0 падающего света из-за рассеяния. Выделим в среде произвольный цилиндр, площадь поперечного сечения которого равна единице, а образующие параллельны оси Z. Вырежем из него бесконечно короткий цилиндрик, ограниченный плоскостями $z = \text{const}$ и $z + dz = \text{const}$. Через первое основание такого цилиндрика ежесекундно вступает энергия $I_0(z)$, а через второе выходит $I_0(z + dz)$. Разность этих энергий $-dI_0$ есть рассеянная энергия $\mathcal{P}_1 N\,dz$.

Приравнивая оба выражения, получим

$$dI_0 = -\gamma I_0 \, dz, \tag{98.15}$$

где в соответствии с формулой (98.12)

$$\gamma = 24\pi^3 \varepsilon_0^2 \left(\frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0} \right)^2 \frac{NV_1^2}{\lambda^4}. \tag{98.16}$$

Таким образом, из-за рассеяния интенсивность падающей волны убывает экспоненциально:

$$I_0 = \mathrm{const} \cdot e^{-\gamma z}. \tag{98.17}$$

Величина γ называется *коэффициентом рассеяния.*

5. Согласно формуле (98.14), впервые полученной Рэлеем, интенсивность рассеянного света обратно пропорциональна *четвертой степени длины волны.* Этот результат называется *законом Рэлея.* Он справедлив для рассеивающих частиц, линейные размеры которых малы по сравнению с длиной волны.

Закон Рэлея качественно подтверждается уже ранними исследованиями Тиндаля, который наблюдал, что белый свет при рассеянии становится синеватым. На основе этих наблюдений Тиндаль высказал мысль, что синий цвет и поляризация неба определяются рассеянием солнечного света на мелких частицах пыли, всегда имеющихся в достаточном количестве в земной атмосфере.

Количественная теория была развита Рэлеем. Если бы рассеяния света не было, то небо было бы совершенно черным. На этом черном фоне звезды и другие небесные светила выделялись бы более ярко и контрастно. Именно таким видят небо космонавты из космических кораблей. При наличии же атмосферы значительная доля прямого солнечного излучения рассеивается в стороны. Она тем больше, чем короче длина волны. Поэтому рассеянный свет обогащен короткими волнами, чем и объясняется *синий цвет неба.* При восходе и заходе Солнца прямой солнечный свет проходит через большую толщу атмосферы, и при этом большая часть коротковолнового излучения теряется на рассеяние. Из прямого света до поверхности Земли доходят преимущественно красные лучи. Вот почему при восходе и заходе Солнце красное. Так же объясняется *красный цвет зари.*

По мере поднятия над земной поверхностью содержание пыли и других посторонних частиц в воздухе уменьшается. Казалось бы, что при этом насыщенность рассеянного света синими лучами должна также уменьшаться. Однако наблюдения в высокогорных обсерваториях показали, что дело обстоит как раз наоборот. Чем чище воздух, чем меньше в нем содержится посторонних частиц, тем насыщеннее излучение неба синими лучами и тем полнее его поляризация. На этом основании Рэлей пришел к заключению, подтвержденному всеми последующими экспериментальными и теоретическими исследованиями, что здесь рассеяние вызывается не посторонними частицами, а *самими молекулами воздуха.* Такое рассеяние света называется *рэлеевским* или *молекулярным рассеянием.* Однако физическая природа молекулярного рассеяния была понята только в 1908 г. М. Смолуховским (1872–1917). *Молекулярное рассеяние вызывается тепловыми флуктуациями показателя преломления*, которые и делают среду оптически мутной. Теория рассеяния света в жидкостях и газах, построенная на этой основе, была создана в 1910 г. Эйнштейном. Она применима в тех случаях, когда длина световой волны настолько велика, что среду можно разбить на объемчики, малые по сравнению с кубом длины волны, каждый из которых содержит, однако, еще очень много молекул. К таким объемчикам еще можно применять макроскопические уравнения Максвелла, не учитывая явно молекулярную структуру вещества. Флуктуации показателя преломления в таких объемчиках и играют роль *макроскопических неоднородностей*, вызывающих рассеяние света, подобно шарикам в предыдущем рассмотрении.

Поле $\mathbf{E}'$, $\mathbf{H}'$ рассеянного излучения в общем случае можно рассчитать с помощью уравнений (98.2). В них теперь ε_0 означает среднее значение диэлектрической проницаемости среды, а $\delta\varepsilon$ — ее флуктуацию. Решение можно получить *методом последовательных приближений.* В нулевом приближении в уравнениях (98.2) пренебрегаем неоднородностями среды, т.е. правыми частями. Тогда рассеянного излучения $\mathbf{E}'$, $\mathbf{H}'$ не будет — останется только падающая волна $\mathbf{E}_0$, $\mathbf{H}_0$. Для нахождения первого приближения в правых частях (98.2) заменяем поле $\mathbf{E}$ его значением $\mathbf{E}_0$ в нулевом приближении. Решая полученные уравнения, находим $\mathbf{E}'$ и $\mathbf{H}'$, а затем $\mathbf{E}$ и $\mathbf{H}$ в первом приближении. Используя полученное решение, уточняем правые части уравнений (98.2) и находим $\mathbf{E}$ и $\mathbf{H}$ во втором приближении, и т.д.

Практически достаточно ограничиться первым приближением, полагая $\delta\mathbf{P} = \delta\varepsilon\mathbf{E}_0/(4\pi)$. Разделим среду на элементарные объемчики $\delta_i V$, малые по сравнению с кубом длины волны, но содержащие еще очень много молекул. Дополнительный дипольный момент объемчика $\delta_i V$, обусловленный флуктуациями диэлектрической проницаемости, будет

$$\mathbf{p} = \frac{\delta_i\varepsilon \cdot \delta_i V}{4\pi}\mathbf{E}_0.$$

Это выражение отличается от (98.6) только коэффициентом при $\mathbf{E}_0$. Поэтому, предполагая падающий свет естественным, для средней интенсивности света, рассеиваемого объемчиком $\delta_i V$, можно сразу написать

$$I_i = \frac{\pi^2}{\lambda^4} \frac{1+\cos^2\theta}{2r^2} I_0 (\delta_i V)^2 \overline{(\delta_i \varepsilon)^2}. \tag{98.18}$$

Соответствующее электрическое поле будет пропорционально $\delta_i V$ и $\delta_i \varepsilon$.

6. Рассмотрим сначала молекулярное рассеяние в *идеальных газах.* В этом случае

$$\varepsilon_i = 1 + 4\pi\beta \frac{N_i}{\delta_i V},$$

где β — поляризуемость молекулы, а N_i — число молекул в объемчике $\delta_i V$. Поскольку объемчик $\delta_i V$ фиксирован в пространстве, т.е. не флуктуирует, из последней формулы получаем $\delta_i V\,\delta\varepsilon_i = 4\pi\beta\delta N_i$, т.е. рассеяние света вызывается *флуктуациями числа молекул* в объемчиках $\delta_i V$. Согласно (97.7), $\overline{(\delta N_i)^2} = \overline{N_i} = N\delta_i V$, где N — число молекул в единице объема. Таким образом,

$$I_i = \frac{8\beta^2\pi^4}{\lambda^4} \frac{1+\cos^2\theta}{r^2} I_0 N \delta_i V.$$

Для газов $n^2 - 1 = 4\pi N\beta$, причем показатель преломления n близок к единице. Поэтому вместо этой формулы можно взять $n - 1 = 2\pi N\beta$. Исключив β, найдем

$$\overline{I}_i = \frac{2\pi^2(n-1)^2}{N\lambda^4} \frac{1+\cos^2\theta}{r^2} I_0 \delta_i V.$$

Чтобы найти интенсивность I света, рассеянного конечным объемом V, заметим, что различные объемчики $\delta_i V$ и $\delta_j V$ рассеивают *некогерентно.* Действительно, возьмем все объемчики $\delta_i V$ одинаковыми по величине. Тогда электрические поля рассеянных ими волн представятся в виде $\mathbf{E}_i = \mathbf{A}\,\delta N_i$, где коэффициент $\mathbf{A}$ одинаков для всех объемчиков. Полное поле рассеянной волны будет $\sum \mathbf{E}_i$, а интенсивность

$$I \sim \overline{\left(\sum \mathbf{E}_i\right)^2} = A^2 \overline{\sum_{i\neq j}\sum \delta N_i\,\delta N_j} + A^2 \overline{\sum (\delta N_i)^2}.$$

Но в идеальных газах числа молекул в различных объемчиках флуктуируют *независимо.* Поэтому двойная сумма в предыдущей формуле обращается в нуль, что и доказывает наше утверждение. Значит, для нахождения I надо $\overline{I}_i$ умножить на число элементарных объемчиков в объеме V, т.е. на $V/\delta_i V$. Это дает

$$I = \frac{2\pi^2(n-1)^2 V}{N\lambda^4} \frac{1+\cos^2\theta}{r^2} I_0. \tag{98.19}$$

Если свет линейно поляризован, то множитель $(1+\cos^2\theta)/2$ следует заменить на $\sin^2\vartheta$.

Формула (98.19) была впервые получена Рэлеем в 1899 г., но недостаточно обоснована им. Рэлей вывел ее в предположении, что рассеяние происходит на *отдельных молекулах газа*, которые ведут себя совершенно аналогично независимым шарикам, о которых шла речь при выводе формулы (98.14). Результирующую интенсивность рассеянного света он вычислял, складывая интенсивности рассеянных волн от отдельных молекул, как если бы эти волны были некогерентны. Он полагал, что некогерентность возникает из-за теплового движения молекул, но не учитывал явно флуктуации числа частиц в рассеивающих объемчиках.

Л.И. Мандельштам в 1907 г. указал, что если бы все (одинаковые) малые пространственно неподвижные объемчики $\delta_i V$ содержали одинаковое число молекул, то излучаемые ими вторичные волны были бы *когерентны, независимо от того, движутся в них молекулы или нет.* При сложении напряженностей полей таких волн происходило бы *интерференционное гашение* их во всех направлениях, за исключением направления падающей волны.

Полного гашения на самом деле не будет, так как (на что в 1908 г. указал Смолуховский) числа молекул в элементарных объемчиках испытывают флуктуации. Из-за таких флуктуаций излучения объемчиков можно разложить на когерентную и некогерентную части. Когерентная часть влияет на падающую волну, изменяя ее фазовую скорость (см. § 68). Некогерентные излучения появляются из-за флуктуаций числа частиц. С ними и связано рассеяние света. Только благодаря случайному соотношению (97.7) окончательный результат (98.19) в случае идеальных газов получается таким же, как если бы молекулы рассеивали свет некогерентно.

7. Формула (98.19) хорошо согласуется с опытом во всех случаях, когда молекулы газа обладают сферической симметрией, т.е. *изотропны.* Таковы, например, молекулы инертных газов. Первое количественное подтверждение эта формула получила при исследовании рассеяния солнечного света в земной атмосфере (синевы неба). Измерив коэффициент рассеяния, можно с помощью формулы (98.19) вычислить число Лошмидта N, т.е. число молекул воздуха в единице объема. Полученное таким путем значение N прекрасно согласуется со значениями этого числа, определенными другими методами.

Если молекулы газа *анизотропны*, например CO_2, то наблюдаются отступления от изложенной теории. Прежде всего, если падающий свет поляризован линейно, то рассеянный свет поляризован только частично, а не полностью, как требует эта теория. Такая *деполяризация рассеянного света* вызывается именно анизотропией молекул. Пусть электрический вектор $\mathbf{E}_0$ падающей волны параллелен оси X. Если бы молекула была изотропна, то ее индуцированный дипольный момент $\mathbf{p} = \beta\mathbf{E}_0$ имел бы то же направление. Свет, рассеянный молекулой, получился бы поляризованным линейно, с плоскостью колебаний, проходящей через ось диполя $\mathbf{p}$ и линию наблюдения. Но если молекула анизотропна, то параллельности между $\mathbf{p}$ и $\mathbf{E}$, вообще говоря, уже не будет. Появятся составляющие вектора $\mathbf{p}$ вдоль осей Y и Z. А так как при тепловом движении ориентация молекулы в пространстве непрерывно и беспорядочно меняется, то поляризуемости молекулы вдоль координатных осей X, Y, Z будут также флуктуировать. Составляющие $\mathbf{p}_y$ и $\mathbf{p}_z$ дают рассеянные волны, поляризация которых отлична от поляризации излучения, даваемого составляющей $\mathbf{p}_x$. Это и приводит к деполяризации рассеянного света.

Флуктуации в ориентации анизотропных молекул вызывают не только деполяризацию, но и влияют на интенсивность рассеянного света. Однако в такие детали теории мы входить не можем.

8. Перейдем теперь к молекулярному рассеянию в *неидеальных газах* и *жидкостях*, когда они находятся в состоянии, не очень близком к критической точке. Исходной будет формула (98.18). Диэлектрическая проницаемость ε является функцией плотности ρ и температуры T среды. От температуры она зависит слабо. Этой зависимостью можно пренебречь. Тогда

$$\delta_i\varepsilon = \frac{d\varepsilon}{d\rho}\,\delta_i\rho,$$

где $\delta_i\rho$ — *флуктуация плотности* в объеме $\delta_i V$. Полагая в формуле (97.25) $V = \delta_i V$, можем написать

$$\delta_i V\overline{(\delta_i\rho)^2} = -\rho^2\frac{kT}{v(\partial P/\partial v)_T}.$$

Следовательно, формула (98.18) переходит в

$$I_i = \frac{\pi^2}{\lambda^4}\frac{1+\cos^2\theta}{2r^2}I_0\delta_i V\left(\rho\frac{d\varepsilon}{\delta\rho}\right)^2\frac{kT}{(-v\,\partial P/\partial v)_T}.$$

Предположим теперь, что рассеивающий объем V не мал. В газах и жидкостях вдали от критической точки линейные размеры элементарных объемчиков $\delta_i V$ можно взять малыми по сравнению с λ, но большими по сравнению с радиусом действия молекулярных сил (порядка 10^{-7} см). В таком случае различные элементарные объемчики будут рассеивать свет *некогерентно*. Для нахождения результирующей интенсивности света, рассеиваемого всем объемом V, надо I_i умножить на число этих объемчиков, т.е. на $V/\delta_i V$. (Мы берем объемчики $\delta_i V$ одинаковой величины, что, очевидно, несущественно и не может влиять на окончательный результат.) Таким путем получаем

$$I = \frac{\pi^2}{\lambda^4}\frac{1+\cos^2\theta}{2r^2} V I_0 \left(\rho\frac{d\varepsilon}{d\rho}\right)^2 \frac{kT}{(-v\,\partial P/\partial v)_T}. \tag{98.20}$$

Эта формула была выведена Эйнштейном в 1910 г. Формула Рэлея (98.19) содержится в ней как частный случай.

Флуктуации плотности становятся очень большими при приближении к *критической точке* рассматриваемого вещества, так как при этом производная $\partial P/\partial v$ стремится к нулю. Этим объясняется так называемая *критическая опалесценция*, т.е. очень сильное рассеяние света в окрестности критической точки. Явление это было известно задолго до создания Смолуховским и Эйнштейном теории флуктуаций, но его причина оставалась неясной вплоть до появления работ этих ученых. В самой критической точке формула Эйнштейна (98.20) дает для интенсивности рассеянного света бесконечное значение. Отсюда следует, что в окрестности критической точки эта формула неприменима. Причина заключается в том, что флуктуации плотности в малых объемчиках $\delta_i V$ в окрестности критической точки уже *нельзя считать статистически независимыми*. Орнштейн и Цернике обобщили формулу Эйнштейна на случай рассеяния света вблизи критической точки. Так как здесь линейные размеры неоднородностей, на которых рассеивается свет, уже не малы по сравнению с λ, то интенсивность рассеянного света не пропорциональна четвертой степени частоты, а возрастает с частотой *более медленно*. Кроме того, рассеяние перестает быть симметричным относительно плоскости, перпендикулярной к падающим лучам: *вперед рассеивается больше света, чем назад*. На всех этих вопросах мы не можем останавливаться.

Формула Эйнштейна учитывает не все причины рассеяния света. В частности, она не учитывает *флуктуации анизотропии* в случае жидкостей, состоящих из анизотропных молекул. Учет этого обстоятельства увеличивает интенсивность рассеянного света и объясняет его деполяризацию. В случае жидкостей соответствующие теоретические расчеты, конечно, не столь надежны, а их согласие с опытом не такое хорошее, как в случае газов.

9. Молекулярное рассеяние света может вызываться не только флуктуациями плотности, но и другими причинами. Интересным примером может служить *рассеяние света в растворах*. В обычных условиях молекулы растворенного вещества настолько равномерно перемешаны с молекулами растворителя, что весь раствор в оптическом отношении представляет собой почти столь же однородную среду, что и чистый растворитель. *Флуктуации концентрации* растворенного вещества в растворителе могут быть причиной рассеяния света. В обычных условиях флуктуации концентрации и обусловленное ими рассеяние света малы. Однако существует много комбинаций веществ, которые ниже определенной температуры $T_к$, называемой *критической температурой смешения*, растворяются друг в друге только частично, а выше этой температуры смешиваются в любых пропорциях (см. т. II, § 123). При критической температуре смешения две жидкости полностью смешиваются друг с другом только *при вполне определенных весовых отношениях*. Такова, например, смесь 20 частей по весу сероуглерода и 80 частей метилового спирта при температуре $T_к = 40$ °С. При критической температуре смешения надо ожидать особенно больших флуктуаций концентрации, подобно тому, что имеет место для флуктуаций плотности вблизи критической точки. Действительно, при критической температуре смешения наблюдается очень сильное рассеяние света, аналогичное критической опалесценции.

10. Вторым примером может служить *рассеяние света на границе раздела двух несмешивающихся жидкостей или на свободной поверхности жидкости*. Из-за теплового движения поверхность жидкости не бывает абсолютно гладкой. Она всегда *неровная*. На этих неровностях свет претерпевает дифракцию, т.е. происходит *поверхностное молекулярное рассеяние*. Если высота неровностей мала по сравнению с длиной волны, как это имеет место в обычных условиях, то амплитуда рассеянного света обратно пропорциональна *первой*, а его интенсивность *второй степени длины волны*. Поверхностное натяжение сглаживает неровности, появившиеся из-за тепловых флуктуаций. Поэтому молекулярное поверхностное рассеяние света наиболее интенсивно на границе двух несмешивающихся жидкостей с близким коэффициентом поверхностного натяжения.

11. Молекулярное рассеяние света в кристаллах впервые было надежно установлено Г.С. Ландсбергом в 1926–1927 гг. Трудность состояла не только в том, что интенсивность рассеянного света в хороших кристаллах по предварительной оценке должна составлять всего около 10^{-8} от интенсивности падающего света. В то время вообще было не ясно, существуют ли кристаллы, в которых основную долю рассеянного света составляет свет *молекулярного рассеяния*, а не паразитный свет, возникающий при рассеянии на различных вкраплениях, микротрещинах и других дефектах кристалла. Метод, с помощью которого удалось отделить одно рассеяние от другого, состоял в исследовании *температурной зависимости интенсивности рассеянного света*. Интенсивность паразитно рассеянного света *не должна зависеть от температуры*, а молекулярно рассеянного — *возрастать с температурой*. Г.С. Ландсберг нашел, что в лучших кристаллах кварца только 25 % рассеянного света не зависит от температуры и, следовательно, вызвано посторонними включениями, а остальные 75 % зависят от температуры линейно, что и указывает на их молекулярное происхождение.

§ 99. Явление Мандельштама–Бриллюэна

1. Как показано в предыдущем параграфе, электромагнитное поле в оптически неоднородной среде с диэлектрической проницаемостью $\varepsilon = \varepsilon_0 + \delta\varepsilon$ может быть представлено в виде $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}'$, $\mathbf{H} = \mathbf{H}_0 + \mathbf{H}'$, где $\mathbf{E}_0$, $\mathbf{H}_0$ — поле падающей, а $\mathbf{E}'$, $\mathbf{H}'$ — рассеянной волн. При слабой неоднородности можно ограничиться линейным приближением и написать

$$\begin{gathered}\operatorname{rot}\mathbf{H}' - \frac{\varepsilon_0}{c}\frac{\partial \mathbf{E}'}{\partial t} = \frac{4\pi}{c}\frac{\partial}{\partial t}\,\delta\mathbf{P}, \quad \operatorname{div}(\varepsilon_0\mathbf{E}') = -4\pi\operatorname{div}(\delta\mathbf{P}),\\ \operatorname{rot}\mathbf{E}' - \frac{1}{c}\frac{\partial \mathbf{H}'}{\partial t} = 0, \quad \operatorname{div}\mathbf{H}' = 0,\end{gathered} \tag{99.1}$$

где

$$\delta\mathbf{P} = \frac{\delta\varepsilon}{4\pi}\,\mathbf{E}_0. \tag{99.2}$$

Эти уравнения показывают, что среда может рассматриваться как *однородная* с диэлектрической проницаемостью ε_0. Влияние фактически имеющихся неоднородностей эквивалентно наличию в среде *дополнительных источников волн*: каждый элемент объема среды dV дает дополнительное излучение как *диполь Герца* с дипольным моментом $\delta\mathbf{P}\,dV$. Это дополнительное излучение и есть *рассеянный свет.*

Уравнения (99.1) *линейны и однородны* как относительно полей $\mathbf{E}'$, $\mathbf{H}'$, так и относительно $\delta\varepsilon$. Отсюда следует, что если представить $\delta\varepsilon$ в виде $\delta\varepsilon = \sum \delta_i\varepsilon$, то в линейном приближении рассеянное излучение может быть получено простой суперпозицией полей, рассеянных на неоднородностях $\delta_i\varepsilon$. Таким образом, можно рассмотреть задачу о рассеянии падающей волны сначала для случая, когда в среде имеется всего *одна неоднородность* $\delta_i\varepsilon$ какого-либо специального вида. При этом неоднородности $\delta_i\varepsilon$, суперпозицией которых представляется $\delta\varepsilon$, можно выбирать произвольно. Рассмотрим сначала случай, когда $\delta\varepsilon$ состоит всего из одного слагаемого $\delta\varepsilon = a\exp(-i\mathbf{K}\mathbf{r})$, где a и $\mathbf{K}$ — постоянные. Пусть падающая волна плоская и представляется выражениями

$$\mathbf{E}_0 = \mathbf{A}e^{i(\omega t - \mathbf{k}\mathbf{r})}, \quad \mathbf{H}_0 = \mathbf{B}e^{i(\omega t - \mathbf{k}\mathbf{r})}.$$

Посмотрим, при каких длинах волн $\lambda = 2\pi/k$ и в каких направлениях будет наблюдаться рассеянное излучение.

Разобьем среду равноотстоящими плоскостями, перпендикулярными к вектору $\mathbf{K}$ (рис. 322). Выберем расстояние между плоскостями равным $\Lambda = 2\pi/K$. Тогда, согласно (99.2), фазы вторичных источников на этих равноотстоящих плоскостях будут одинаковы. Если бы неоднородность была только в слое I, а дальше среда была однородна, то падающая волна претерпела бы отражение от этого слоя и частично прошла бы дальше. При наличии неоднородности только в слое II мы получили бы другую отраженную волну с той же амплитудой, но иной фазой. При наличии неоднородности в слое III получилась бы третья отраженная волна, и т.д. В линейном приближении поле рассеяния всей среды равно простой суперпозиции этих отраженных волн. Чтобы они не гасили, а усиливали друг друга, необходимо выполнение *условия Брэгга–Вульфа*: $2\Lambda\sin(\theta/2) = m\lambda$, где θ — угол рассеяния, т.е. угол между направлениями падающего и рассеянного излучений, а m — целое число (порядок дифракционного спектра).

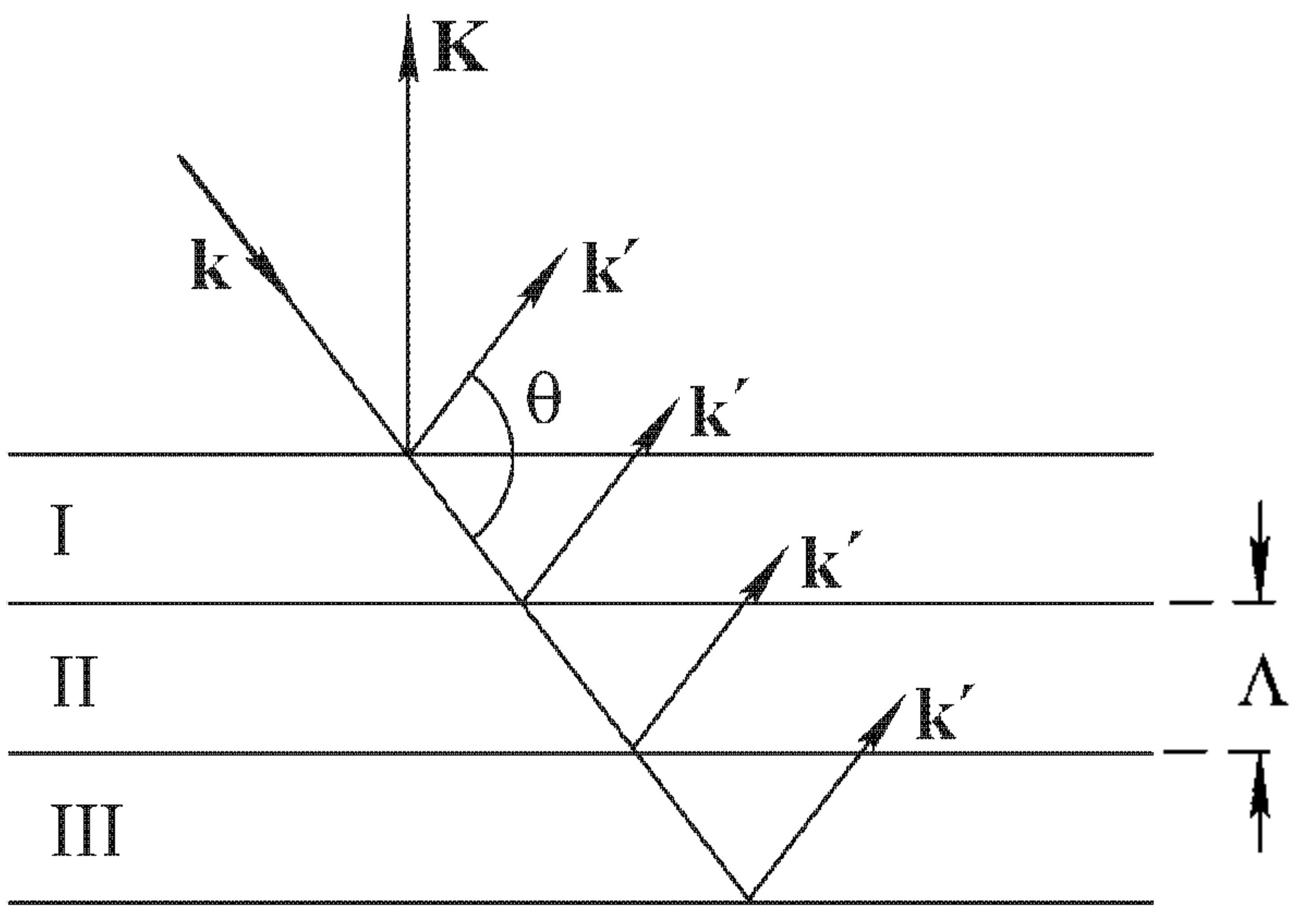


Рис. 322

Покажем, что $m = 1$. Все плоские волны, отраженные различными слоями, складываясь, дают волну вида $\mathbf{E}' = \mathbf{A}'e^{i(\omega t - \mathbf{k}'\mathbf{r}')}$, где волновой вектор $\mathbf{k}'$ определяет направление распространения отраженных волн. С другой стороны, дополнительная поляризация среды

$$\delta\mathbf{P} = \frac{\mathbf{E}_0}{4\pi}\,\delta\varepsilon = \frac{a\mathbf{A}}{4\pi}\,e^{i[\omega t - (\mathbf{k}+\mathbf{K})\mathbf{r}]}.$$

Подставляя эти выражения во второе уравнение (99.1) и сравнивая показатели, легко получить $\mathbf{k}' - \mathbf{k} = \mathbf{K}$, откуда

$$2\Lambda\sin(\theta/2) = \lambda. \tag{99.3}$$

Таким образом, при дифракции волны на синусоидальной неоднородности диэлектрической проницаемости в линейном приближении получается дифракционный спектр *только первого порядка.*

Любую неоднородность в среде можно по теореме Фурье представить *в виде суперпозиции плоских синусоидальных неоднородностей различных направлений.* Согласно доказанному выше такие синусоидальные неоднородности рассеивают свет *независимо* друг от друга. Но при фиксированном направлении рассеянного излучения эффективны не все синусоидальные неоднородности, а только такие, волновой вектор $\mathbf{K}$ которых направлен *по биссектрисе угла, дополнительного к θ* до 180° (рис. 322). Остальные синусоидальные неоднородности для рассеяния в рассматриваемом направлении не играют роли. Мы видим, что механизм рассеяния света на неоднородностях диэлектрической проницаемости вполне аналогичен механизму рассеяния рентгеновских лучей в кристаллах в той форме, в какой он был представлен Вульфом и Брэггом (см. § 61).

2. До сих пор мы принимали во внимание изменения функции $\delta\varepsilon$ *в пространстве*, но не учитывали ее изменения *во времени*. Учет последнего обстоятельства приводит к новому явлению в рассеянии света. Считая, как и в предыдущем параграфе, ε функцией только плотности ρ, напишем в линейном приближении $\Delta\varepsilon = (d\varepsilon/d\rho)\Delta\rho$. Всякая неоднородность плотности, возникшая в среде, является *источником звуковых волн*. Разложим $\Delta\rho$ в интеграл или ряд Фурье и возьмем в этом разложении только те звуковые волны, которые существенны для рассеяния волн в рассматриваемом направлении. Их волновой вектор $\mathbf{K}$ был определен выше. Этому значению $\mathbf{K}$ соответствует определенная звуковая частота Ω и два направления распространения звуковой волны: вдоль $\mathbf{K}$ и против $\mathbf{K}$. Неоднородность $\delta\varepsilon$, вызывающая рассеяние света в рассматриваемом направлении, представится суммой $\delta\varepsilon = \delta\varepsilon_1 + \delta\varepsilon_2$, где $\delta\varepsilon_1$ и $\delta\varepsilon_2$ имеют вид плоских звуковых волн:

$$\delta\varepsilon_1 = a_1 e^{i(\Omega t - \mathbf{Kr})} \quad \text{и} \quad \delta\varepsilon_2 = a_2 e^{-i(\Omega t + \mathbf{Kr})}.$$

Им соответствуют векторы дополнительной поляризации среды:

$$\delta\mathbf{P}_1 = \frac{\mathbf{E}_0}{4\pi}\,\delta\varepsilon_1 = \frac{a_1\mathbf{A}}{4\pi}\,e^{i[(\omega+\Omega)t-(\mathbf{k}+\mathbf{K})\mathbf{r}]},$$

$$\delta\mathbf{P}_2 = \frac{a_2\mathbf{A}}{4\pi}\,e^{i[(\omega-\Omega)t-(\mathbf{k}+\mathbf{K})\mathbf{r}]}.$$

Таким образом, источники рассеянного излучения, а значит и само рассеянное излучение, будут меняться во времени с частотами $\omega+\Omega$ и $\omega-\Omega$ (модуляция световой волны акустической волной). В спектре рассеянного излучения должен наблюдаться дублет с теми же частотами. Это явление называется *тонкой структурой линий рэлеевского рассеяния* или *рассеянием Мандельштама–Бриллюэна*. Смещение частоты равно $\Omega = Kv = (2\pi/\Lambda)v$, где v — скорость звука, а Λ — длина звуковой волны. На основании (99.3)

$$\Omega = \frac{4\pi v}{\lambda}\sin\frac{\theta}{2} = 2\omega n\frac{v}{c}\sin\frac{\theta}{2}, \tag{99.4}$$

где c — скорость света в вакууме, а n — показатель преломления среды.

Дублет Мандельштама–Бриллюэна можно трактовать как *доплеровское изменение частоты света* при отражении от акустической волны. Когда акустическая волна распространяется навстречу световой, происходит увеличение частоты света, в противоположном случае — уменьшение. Доплеровское изменение частоты определяется формулой

$$\frac{\Omega}{\omega} = \frac{2v \sin\theta/2}{c/n},$$

откуда и получается формула (99.4).

Представление о тепловом движении как о звуковых волнах всевозможных частот и направлений распространения было введено Дебаем (1884–1966) в его теории теплоемкости твердых тел. К ним Дебай применял методы статистической физики. Это — те же волны, которые вызывают рассеяние света и дублет Мандельштама–Бриллюэна.

3. Тонкая структура линий рэлеевского рассеяния была предсказана независимо друг от друга Л.И. Мандельштамом и Л. Бриллюэном. По свидетельству Г.С. Ландсберга, Л.И. Мандельштам выполнил свою работу еще в 1918 г., хотя краткая заметка о ней появилась значительно позже, в 1926 г., когда часть найденных Д.И. Мандельштамом результатов была уже опубликована Бриллюэном (1922 г.). Мандельштам и Ландсберг пытались на опыте обнаружить предсказанное явление при рассеянии света в кварце. Качественно им удалось констатировать существование явления. Однако недостаточная разрешающая способность их спектральной аппаратуры не позволяла исследовать его количественно. Кроме того, эти опыты привели их к открытию *комбинационного рассеяния света* (см. § 100). Естественно, что их внимание переключилось на исследование этого более важного явления. По их предложению исследованием тонкой структуры рэлеевского рассеяния занялся Е.Ф. Гросс (1897–1972) в Ленинграде.

Гросс обнаружил явление при рассеянии света *в жидкостях* [1]. Однако оказалось, что в жидкостях, наряду с двумя смещенными компонентами, наблюдается также и *несмещенная компонента.* Происхождение несмещенной компоненты было объяснено Ландау (1908–1968) и Плачеком (1905–1955).

---

[1] Это было неожиданно, так как, согласно гидродинамической теории, поглощение звука в жидкостях пропорционально квадрату частоты ω. Если бы гидродинамическая теория была верна без ограничений, то звуковые волны оптических частот в жидкостях распространяться не могли бы. Обнаружение тонкой структуры в жидкостях послужило поводом Л.И. Мандельштаму и М.А. Леонтовичу (р. 1903) к разработке релаксационной молекулярной теории вязкости жидкостей и основанной на ней теории поглощения звука.

Рассматривая удельный объем жидкости V как функцию давления и энтропии, можно написать

$$\delta V = (\partial V/\partial P)_S\, \delta P + (\partial V/\partial S)_P\, \delta S. \qquad (99.5)$$

Отсюда видно, что существует два вида флуктуаций удельного объема: одни вызваны *флуктуациями давления* при постоянной энтропии, другие — *флуктуациями энтропии* при постоянном давлении. Флуктуации первого типа распространяются в виде *акустических волн* и ведут к появлению *смещенных компонент*. Флуктуационные неоднородности второго типа выравниваются посредством *теплопроводности*, а следовательно, распространяются значительно более медленно, — они и ведут к появлению в рассеянном свете *несмещенной компоненты*.

Для количественного исследования заметим, что процессы рассеяния света на флуктуациях давления и энтропии *некогерентны*. Поэтому интегральные интенсивности несмещенной I_ω и смещенных $I_{\omega-\delta\omega}$, $I_{\omega+\delta\omega}$ компонент связаны соотношением

$$\frac{I_\omega}{I_{\omega-\delta\omega}+I_{\omega+\delta\omega}} = \frac{(\partial V/\partial s)_P^2\,\overline{\Delta s^2}}{(\partial V/\partial P)_s^2\,\overline{\Delta P^2}} = -\left(\frac{\partial V}{\partial s}\right)_P^2 \left(\frac{\partial s}{\partial T}\right)_P \left(\frac{\partial P}{\partial V}\right)_s,$$

где использованы выражения (97.22) и (97.24), а также формула $c_P = T(\partial s/\partial T)_P$, причем малой буквой s обозначена удельная энтропия. Так как дифференциал удельной энтальпии $di = T\,ds + V\,dP$ — полный дифференциал, то $(\partial T/\partial P)_s = (\partial V/\partial s)_P$. Поэтому

$$\frac{I_\omega}{I_{\omega-\delta\omega}+I_{\omega+\delta\omega}} = -\left(\frac{\partial T}{\partial P}\right)_s \left(\frac{\partial P}{\partial V}\right)_s \left(\frac{\partial V}{\partial s}\right)_P \left(\frac{\partial s}{\partial T}\right)_P = -\left(\frac{\partial T}{\partial V}\right)_s \left(\frac{\partial V}{\partial T}\right)_P,$$

или, на основании тождества $(\partial T/\partial V)_s(\partial V/\partial s)_T(\partial s/\partial T)_V = -1$,

$$\frac{I_\omega}{I_{\omega-\delta\omega}+I_{\omega+\delta\omega}} = \left(\frac{\partial V}{\partial T}\right)_P \left(\frac{\partial s}{\partial V}\right)_T \left(\frac{\partial T}{\partial s}\right)_V = \frac{(\partial V/\partial T)_P(\partial s/\partial V)_T}{(\partial s/\partial T)_V}.$$

Рассматривая энтропию s как функцию T и V, получим

$$\left(\frac{\partial s}{\partial T}\right)_P = \left(\frac{\partial s}{\partial T}\right)_V + \left(\frac{\partial s}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_P.$$

Окончательно:

$$\frac{I_\omega}{I_{\omega-\delta\omega}+I_{\omega+\delta\omega}} = \frac{(\partial s/\partial T)_P - (\partial s/\partial T)_V}{(\partial s/\partial T)_V} = \frac{c_P - c_v}{c_v}. \qquad (99.6)$$

Эта формула была получена Ландау и Плачеком.

В аморфных твердых телах звуковые волны могут быть *продольными* и *поперечными*. Они распространяются с различными скоростями. Поэтому в рассеянном свете спектральная линия должна расщепляться на *пять компонент*: одну несмещенную и две пары смещенных компонент, из которых одна пара получается от рассеяния на продольных акустических волнах, а другая — на поперечных.

В.В. Владимирский (р. 1915) указал, что в кристаллах в общем случае спектральная линия неполяризованного света должна расщепляться на 25 *компонент*: одну несмещенную и 24 смещенные. Дело в том, что в кристалле в каждом направлении могут распространяться одна продольная акустическая волна и две поперечные. В том же направлении могут распространяться две световые волны, поляризованные во взаимно перпендикулярных плоскостях. Каждая из этих световых волн в свою очередь расщепляется на две волны при отражении от акустических волн соответствующих направлений распространения. Это и приводит к появлению в рассеянном свете 24 смещенных компонент. Однако из-за слабой анизотропии всех исследованных кристаллов эти 24 компоненты обычно группируются в шесть групп по четыре линии в каждой и не разрешаются спектральными приборами. На опыте наблюдаются *шесть смещенных компонент.*

4. С изобретением лазеров стала возможной генерация мощных (так называемых гигантских) световых импульсов, оказывающих существенное воздействие на среду, в которой распространяется свет. В переменном электрическом поле $\mathbf{E}$ возникает электрострикционное давление

$$\mathcal{P} = \frac{1}{8\pi}\left(\rho\frac{d\varepsilon}{d\rho}\right)E^2 \tag{99.7}$$

(см. т. III, § 32). Величина $\rho\, d\varepsilon/d\rho$ порядка единицы. В слабых световых полях, с которыми имеет дело линейная оптика, давление $\mathcal{P}$ ничтожно и его влиянием на среду можно полностью пренебречь. Но в световом поле гигантского лазерного импульса это давление может достигать сотни тысяч атмосфер. Тогда световые и акустические волны в среде надо рассматривать *совместно.* Они описываются сложной системой взаимосвязанных *нелинейных уравнений* электродинамики и акустики. Это приводит к ряду *нелинейных оптических явлений.* Одним из них является *вынужденное рассеяние Мандельштама–Бриллюэна.* Хотя нелинейные оптические явления будут разбираться в главе XI, возникновение вынужденного рассеяния Мандельштама–Бриллюэна удобнее разобрать уже здесь.

Пусть $\mathbf{E}_0 = \mathbf{A}_0 \cos(\omega t - \mathbf{kr})$, $\mathbf{E}_1 = \mathbf{A}_1 \cos[(\omega + \Omega)t - \mathbf{k}'\mathbf{r}]$, $\mathbf{E}_2 = \mathbf{A}_2 \cos[(\omega - \Omega)t - \mathbf{k}'\mathbf{r}]$, $\mathbf{E}_3 = \mathbf{A}_3 \cos[\omega t - \mathbf{k}'\mathbf{r}]$ — напряженности электрического поля падающей и трех рассеянных волн Мандельштама–Бриллюэна. Последние три волны возникают при рассеянии на тепловых флуктуациях. Интенсивности их сначала малы, но в дальнейшем могут усилиться за счет взаимодействия с падающей волной. Электрострикционное давление $\mathcal{P}$ определяется квадратом суммы всех полей, т.е. $(\mathbf{E}_0 + \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3)^2$. При возведении в квадрат представим по известным формулам тригонометрии квадраты и произведения косинусов в виде сумм постоянных членов и косинусов суммарных и разностных аргументов. Постоянные члены для возбуждения звуковых волн не играют роли. Не имеют значения и члены с косинусами от суммарных аргументов. Это — *высокочастотные члены*, меняющиеся во времени с оптическими частотами, а звуковые волны быстро затухают с увеличением частоты. Возбуждение звуковых волн связано только с *низкочастотными членами*, содержащими косинусы разностных аргументов. Выпишем все эти члены, опуская при этом числовые коэффициенты и принимая во внимание соотношение $\mathbf{K} = \mathbf{k}' - \mathbf{k}$ (см. рис. 322). Получим

$$\mathbf{A}_0\mathbf{A}_1 \cos(\Omega t - \mathbf{Kr}), \quad \mathbf{A}_0\mathbf{A}_2 \cos(\Omega t + \mathbf{Kr}), \quad \mathbf{A}_0\mathbf{A}_3 \cos \mathbf{Kr},$$

$$\mathbf{A}_1\mathbf{A}_2 \cos 2\Omega t, \quad \mathbf{A}_1\mathbf{A}_3 \cos \Omega t, \quad \mathbf{A}_2\mathbf{A}_3 \cos \Omega t.$$

Из этих членов имеет значение *только первый*. Он представляет волну, распространяющуюся в том же направлении и с той же фазой, что и первичная звуковая волна, возникшая из-за тепловых флуктуаций. Поэтому будет происходить *параметрическое усиление* этой акустической волны и всех световых волн, рассеянных на ней. Такой процесс усиления будет продолжаться до тех пор, пока интенсивность рассеянного света не станет сравнимой с интенсивностью падающего. Это действительно и наблюдается на опыте. В отличие от некогерентного рассеяния на тепловых флуктуациях, *вынужденное рассеяние Мандельштама–Бриллюэна когерентно.*

ЗАДАЧА

При рассеянии света резонансной линии ртутной лампы ($\lambda = 253{,}65$ нм) в кристалле алмаза под углом $\theta = 90^\circ$ к направлению падающего пучка были найдены две пары смещенных компонент с $\delta\lambda = 0{,}052$ нм и $\delta\lambda = 0{,}032$ нм. (Речь идет о смещении относительно центральной — несмещенной — компоненты.) Определить скорости продольной и поперечной акустических волн в алмазе. Показатель преломления алмаза $n = 2{,}42$.

Ответ. $v = \dfrac{c}{2n\sin(\theta/2)}\dfrac{\delta\lambda}{\lambda}$; $v_{\text{прод}} = 18\,000$ м/с; $v_{\text{попереч}} = 11\,000$ м/с.

Благодаря большой скорости звука в алмазе тонкую структуру линий рэлеевского рассеяния удается исследовать даже с помощью призменных спектрографов.

§ 100. Комбинационное рассеяние света

1. При спектральных исследованиях рассеяния света в кварце и исландском шпате (февраль 1928 г.) Мандельштам и Ландсберг обнаружили, что каждая спектральная линия падающего света сопровождается появлением системы линий измененной частоты, называемых *сателлитами* (спутниками). Практически одновременно то же явление было открыто в Индии Раманом (1888–1970) и Кришнаном (1898–1961) при исследовании рассеяния света в жидкостях. Изменение длины волны оказалось значительно больше, чем при рассеянии Мандельштама–Бриллюэна. Это явление называется *комбинационным рассеянием света* или *эффектом Рамана* (который раньше послал в печать сообщение о своем открытии). Приведем основные законы комбинационного рассеяния, установленные экспериментально.

1) Частоты сателлитов отличаются от частоты возбуждающей их линии на $\Delta\omega^j_{\text{комб}}$, где j — номер сателлита, так что различным сателлитам соответствуют различные $\Delta\omega_{\text{комб}}$. При переходе от одной спектральной линии первичного пучка к другой совокупность значений $\Delta\omega^j_{\text{комб}}$ остается *одной и той же*. Она характерна для рассматриваемого вещества и меняется только при переходе от одного вещества к другому. Это обстоятельство используется в спектральном анализе методом комбинационного рассеяния света. Обычно частоты спектральных линий и их изменения принято характеризовать числом волн, укладывающихся в 1 см. Их мы будем обозначать через ν и $\Delta\nu$, полагая $\nu = 1/\lambda$. Таблица 11 показывает, насколько хорошо по измерени-

Таблица 11

Частота основной линии ν, см$^{-1}$	Бензол C_6H_6	Хлороформ $CHCl_3$		Четыреххлористый углерод CCl_4	
	$\Delta\nu_{\text{красн}}$	$\Delta\nu_{\text{красн}}$	$\Delta\nu_{\text{фиол}}$	$\Delta\nu_{\text{красн}}$	$\Delta\nu_{\text{фиол}}$
33 692	992,9	—	—	—	—
33 086	991,3	667,7	—	459,7	—
31 985	992,9	667,7	669,3	458,3	459,4
31 923	992,3	667,3	669,5	457,8	460,4
29 917	—	668,1	670,5	458,7	458,8
27 388	991,2	666,6	—	458,4	461,2
27 354	991,3	666,8	669,9	459,4	461,8
27 290	991,1	—	669,8	459,1	461,8
24 705	992,2	669,2	669,5	457,9	459,7
24 515	—	667,5	—	457,8	—
23 034	992,9	667,3	—	—	—
22 994	992,1	670,8	—	457,8	—
22 938	992,3	669,4	667,1	459,6	457,8
Среднее	992,0	668,0	669,4	458,5	460,2

ям Г.С. Ландсберга соблюдается постоянство $\Delta\nu$ для различных длин волн в комбинационном рассеянии света.

2) Каждому сателлиту с частотой $\omega - \Delta\omega_{\text{комб}}$, смещенной в красную сторону спектра, соответствует сателлит с частотой $\omega + \Delta\omega_{\text{комб}}$, смещенной в фиолетовую сторону. Первые сателлиты называются *красными* или *стоксовыми*, вторые — *фиолетовыми* или *антистоксовыми*.

3) Число различных сателлитов (т.е. число значений индекса j) и их относительная интенсивность при одной и той же температуре также зависят от рассеивающего вещества. Но *интенсивности фиолетовых сателлитов значительно меньше интенсивностей соответствующих им красных сателлитов и проявляют общую тенденцию к ослаблению по мере увеличения* $\Delta\nu_{\text{комб}}$.

4) Постоянные $\Delta\omega_{\text{комб}}$, *характерные для рассматриваемого вещества, совпадают с собственными частотами* $\Omega_{\text{инфр}}$ *инфракрасных колебаний того же вещества*, хотя не всем известным $\Omega_{\text{инфр}}$ находятся соответствующие $\Delta\omega_{\text{комб}}$, и наоборот. Кроме того, нередко интенсивной линии комбинационного рассеяния соответствует слабая линия инфракрасного поглощения и наоборот.

5) Линии комбинационного рассеяния света более или менее *поляризованы*. Степень поляризации различных сателлитов одной и той же линии не одинакова и не находится в прямой связи с поляризацией основной линии рассеянного света. Характер поляризации красных и фиолетовых сателлитов, соответствующих данному значению $\Delta\omega_{\text{комб}}$, *всегда одинаков и не зависит от частоты основной линии*.

2. Явление комбинационного рассеяния света было объяснено сразу же Мандельштамом и Ландсбергом, когда они открыли это явление.

В поле световой волны $\mathbf{E}$ электроны внутри молекулы приходят в колебания, и молекула приобретает индуцированный дипольный момент $\mathbf{p} = \beta\mathbf{E}$. С классической точки зрения тензор поляризуемости молекулы β определяется мгновенными положениями ее атомных ядер. Но сами ядра не находятся в покое, а совершают беспорядочное тепловое движение. По этой причине поляризуемость β не остается постоянной, а меняется во времени. Ее можно представить наложением гармонических колебаний, частоты которых определяются колебаниями атомных ядер, т.е. совпадают с собственными частотами инфракрасных колебаний молекулы. Возникает *модуляция колебаний* индуцированных дипольных моментов $\mathbf{p}$.

Если внешнее электрическое поле $\mathbf{E}$ меняется во времени гармонически с частотой ω, то в колебаниях дипольного момента $\mathbf{p}$ появятся комбинационные частоты $\omega \pm \Omega_{\text{инфр}}$. Такие же частоты появятся и в излучении этих дипольных моментов, т.е. в рассеянном свете. Таково с классической точки зрения происхождение комбинационного рассеяния.

Изложенные соображения можно облечь в математическую форму. Если число ядер в молекуле равно s, то они обладают $3s$ степенями свободы. Из них три степени свободы *поступательные*, а три *вращательные*. Остальные $f = 3s - 6$ степеней свободы приходятся на *внутреннее движение ядер молекулы*. Для описания внутреннего движения ядер требуется f координат $q_1, q_2, \ldots, q_j$. Выбор координат произволен. Удобнее всего взять *нормальные обобщенные координаты*.

В положении равновесия ядер все координаты равны нулю. При малых отклонениях из положений равновесия каждая координата q_m при тепловом движении совершает свободное гармоническое колебание $q_m = a_m \cos(\Omega_m t + \delta_m)$ с инфракрасной частотой Ω_m и хаотически меняющейся фазой δ_m. В силу малости колебаний тензор β можно разложить в ряд и ограничиться первыми степенями по q. Считая для простоты β скаляром, получим

$$\beta = \beta_0 + \sum \left(\frac{\partial\beta}{\partial q_m}\right) q_m,$$

или после подстановки значений q_m

$$\beta = \beta_0 + \sum \left(\frac{\partial\beta}{\partial q_m}\right) a_m \cos(\Omega_m t + \delta_m) =$$
$$= \beta_0 + \frac{1}{2}\sum \left(\frac{\partial\beta}{\partial q_m}\right) a_m e^{i(\Omega_m t+\delta_m)} + \frac{1}{2}\sum \left(\frac{\partial\beta}{\partial q_m}\right) a_m e^{-i(\Omega_m t+\delta_m)}.$$

Падающую волну запишем в комплексной форме $\mathbf{E} = \mathbf{E}_0 e^{i\omega t}$. Тогда

$$\mathbf{p} = \beta_0 \mathbf{E}_0 e^{i\omega t} +$$
$$+ \frac{\mathbf{E}_0}{2}\sum \left(\frac{\partial\beta}{\partial q_m}\right) a_m e^{i[(\omega+\Omega_m)t+\delta_m]} + \frac{\mathbf{E}_0}{2}\sum \left(\frac{\partial\beta}{\partial q_m}\right) a_m e^{i[(\omega-\Omega_m)t-\delta_m]}.$$

Отсюда видно, что в рассеянном излучении появляется свет *не только с исходной частотой* ω, *но и с комбинированными частотами* $\omega \pm \Omega_m$. Ясно, что волны, рассеиваемые отдельными молекулами, *некогерентны*, так как при тепловом возбуждении колебаний ядер фазы δ_m меняются нерегулярно при переходе от одной молекулы к другой и от одного колебания к другому.

3. Классическая теория комбинационного рассеяния, как и всякого явления, связанного с излучением и поглощением света, конечно, недостаточна. Успехи этой теории связаны с *относительно большими* массами атомных ядер, благодаря чему многие особенности их колебаний правильно передаются уравнениями классической механики. Однако классическая теория объясняет не все закономерности комбинационного рассеяния света. В частности, она не может объяснить соотношения между интенсивностями соответственных красных и фиолетовых сателлитов. По классической теории эти интенсивности должны быть *практически одинаковы*. Опыт же показывает, что *интенсивность красных сателлитов всегда больше интенсивности соответственных фиолетовых сателлитов.*

Квантовая теория дает естественное объяснение этой закономерности. Согласно квантовой механике энергетические уровни молекулы образуют *дискретный ряд* $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \ldots$ Рассеяние фотона на молекуле аналогично процессу столкновения его с молекулой, к которому применим *закон сохранения энергии.* В таком процессе фотон может либо передать часть своей энергии молекуле, либо, наоборот, получить энергию от возбужденной молекулы. Согласно соотношению $\mathcal{E}_{\text{фот}} = \hbar\omega$, изменение энергии фотона проявляется в *изменении его частоты.* Пусть фотон с энергией $\hbar\omega$ рассеивается на молекуле с энергией $\mathcal{E}_n$. В результате получается рассеянный фотон с энергией $\hbar\omega'$, а молекула переходит на энергетический уровень $\mathcal{E}_m$. По закону сохранения энергии $\hbar\omega + \mathcal{E}_n = \hbar\omega' + \mathcal{E}_m$, откуда

$$\omega' = \omega + \Omega_{nm}, \tag{100.1}$$

где

$$\Omega_{nm} = (\mathcal{E}_n - \mathcal{E}_m)/\hbar. \tag{100.2}$$

Если $\mathcal{E}_n > \mathcal{E}_m$, то частота рассеянного фотона больше, чем падающего, т.е. в рассеянном свете появляется *фиолетовый сателлит.* Если же $\mathcal{E}_n < \mathcal{E}_m$, то при рассеянии возникает *красный сателлит.*

Пусть $\mathcal{E}_n < \mathcal{E}_m$. Тогда красный сателлит появится, когда исходным уровнем является уровень $\mathcal{E}_n$, а фиолетовый — когда исходным будет уровень $\mathcal{E}_m$. Для отношения интенсивностей сателлитов можно написать $I_{\text{кр}}/I_{\text{фиол}} = N_n/N_m$, где N_n — число молекул на уровне $\mathcal{E}_n$, а N_m — на уровне $\mathcal{E}_m$. Согласно формуле Больцмана, при тепловом равновесии

$$\frac{N_n}{N_m} = \exp\left(-\frac{\mathcal{E}_n - \mathcal{E}_m}{kT}\right) = \exp\frac{\hbar|\Omega_{nm}|}{kT},$$

а потому

$$\frac{I_{\text{кр}}}{I_{\text{фиол}}} = \exp\frac{\hbar|\Omega_{nm}|}{kT}. \tag{100.3}$$

Эта формула вполне объясняет наблюдаемое соотношение интенсивностей.

4. В мощных импульсах лазерного излучения наблюдается нелинейное явление, называемое *вынужденным комбинационным рассеянием света.* Оно возникает из-за *обратного воздействия световой волны на молекулы среды.* В неоднородном электрическом поле $\mathbf{E}$ на молекулу с дипольным моментом $\mathbf{p}$ действует сила $\mathbf{F} = (\mathbf{p}\nabla)\mathbf{E}$. Силы такого рода действуют и на *части молекулы*, так как всякая электрически нейтральная часть молекулы, состоящая, например, из ядра и электрона, обладает дипольным моментом. Индуцированные дипольные моменты $\mathbf{p}$ пропорциональны полю $\mathbf{E}$, так что все эти силы *квадратично* зависят от поля. Поле $\mathbf{E}$ слагается из поля падающей волны $\mathbf{E}_0$ и поля рассеянных волн $\mathbf{E}'$. Первоначально поле $\mathbf{E}'$ слабое, так как оно возникает из-за тепловых флуктуаций в среде. Но затем оно может усиливаться из-за взаимодействия с падающей волной. Среди слагающих сил

$$[(\mathbf{E}_0 + \mathbf{E}')\nabla](\mathbf{E}_0 + \mathbf{E}')$$

присутствуют члены с произведением полей $\mathbf{E}_0$ и $\mathbf{E}'$, частоты которых совпадают с соответствующими частотами инфракрасных колебаний молекулы. Они вызовут *резонансное усиление* таких инфракрасных колебаний, что поведет к увеличению интенсивности соответствующих линий комбинационного рассеяния. Вынужденные колебания ядер молекулы происходят в фазе с падающей волной, а потому, в отличие от обычного (теплового) комбинационного рассеяния, *вынужденное комбинационное рассеяние когерентно с падающей волной.*

ГЛАВА IX

ТЕОРИЯ ОТНОСИТЕЛЬНОСТИ

§ 101. Принцип относительности в ньютоновской и релятивистской механике

1. *Теория относительности — фундаментальная физическая теория, охватывающая всю физику.* Она возникла в начале XX века в результате преодоления принципиальных трудностей, с которыми столкнулась электродинамика и оптика движущихся тел. Настойчивые и мучительные попытки преодолеть эти трудности на основе гипотезы мирового эфира окончились неудачей. Теория относительности отказалась от использования гипотезы эфира. В ее основе лежат не гипотезы, а *принципы* или *постулаты*, твердо установленные экспериментально. В этом сила теории относительности, причина ее успехов. В нашу задачу не входит подробное изложение истории теории относительности и экспериментального обоснования ее постулатов. Мы коснемся этих вопросов предельно кратко, чтобы лучше оттенить принципиальные моменты, глубже уяснить смысл постулатов и содержание теории относительности.

Творцом теории относительности является Альберт Эйнштейн. Теория относительности была изложена им в 1905 г. в основополагающей работе «К электродинамике движущихся тел». Многие результаты этой работы были получены ранее Лармором (1857–1942), Лорентцом и Пуанкаре (1854–1912). Однако и Лармор и Лорентц принципиально стояли на точке зрения неподвижного эфира, с которым связывалась преимущественная система отсчета. Ближе всего к теории относительности был Пуанкаре, который еще в 1898 г. дал критику понятия одновременности пространственно разделенных событий, повторенную в дальнейшем Эйнштейном. Пуанкаре заполнил также математические пробелы и устранил ошибки, допущенные Лорентцом. Однако принципиально новое и глубокое физическое понимание всей проблемы и последовательное построение теории относительности с единой точки зрения содержится лишь в упомянутой выше работе Эйнштейна, написанной к тому же без всякого влияния своих предшественников.

2. Начнем с повторения того, что было сказано в первом томе о принципе относительности в ньютоновской механике. Основной закон ньютоновской механики выражается уравнением

$$m\ddot{\mathbf{r}} = \mathbf{F}, \tag{101.1}$$

где $\mathbf{F}$ — сила, действующая на материальную точку, а $\mathbf{r}$ — радиус-вектор, определяющий положение последней относительно какой-либо инерциальной системы отсчета. Возьмем две инерциальные системы отсчета, одну из которых будем обозначать через S и называть *неподвижной* (или *нештрихованной*), а другую — через S', называя ее *движущейся* (или *штрихованной*). Пусть система S' движется относительно системы S равномерно и прямолинейно со скоростью $\mathbf{V}$. Если $\mathbf{r}$ и $\mathbf{r}'$ — радиусы-векторы, определяющие положения движущейся материальной точки относительно этих систем отсчета в момент времени t, то они связаны между собой *преобразованием Галилея*

$$\mathbf{r} = \mathbf{r}' + \mathbf{V}t. \tag{101.2}$$

При этом в ньютоновской механике предполагается, что время t *абсолютно*, т.е. одинаково во всех системах отсчета. Для простоты отсчет времени ведется с того момента, когда начала координат систем S и S' совмещаются между собой. Это не отражается на общности рассуждений.

Дважды дифференцируя соотношение (101.2) по времени, находим *формулы преобразования скорости и ускорения*:

$$\mathbf{v} = \mathbf{v}' + \mathbf{V}, \quad \mathbf{a} = \mathbf{a}'. \tag{101.3}$$

Таким образом, *ускорение инвариантно относительно преобразования Галилея*, т.е. одинаково в обеих системах отсчета S и S'. Радиусы-векторы $\mathbf{r}$ и $\mathbf{r}'$, скорости $\mathbf{v}$ и $\mathbf{v}'$ не одинаковы. Однако *разности радиусов-векторов, а также скоростей двух любых материальных точек одни и те же*, поскольку они определяют *относительные положения и скорость* одной точки относительно другой. Сила $\mathbf{F}$ в ньютоновской механике зависит только от *разностей радиусов-векторов и скоростей* взаимодействующих материальных точек. Поэтому она, а с ней и уравнение Ньютона (101.1) не меняются при преобразовании Галилея. То же относится и к дифференциальным уравнениям движения систем материальных точек в механике Ньютона. Таким образом, получается следующий результат:

Уравнения механики Ньютона, определяющие изменения состояния движения механических систем, инвариантны относительно преобразования Галилея.

Это положение называется *принципом относительности Галилея.* Ему можно дать также следующую формулировку:

Законы природы, определяющие изменения состояния движения механических систем, не зависят от того, к какой из двух инерциальных систем отсчета, движущихся одна относительно другой прямолинейно и равномерно, они относятся.

Дорелятивистская физика считала обе формулировки тождественными, поскольку при равномерном поступательном движении инерциальных систем отсчета относительно друг друга она не допускала никакого другого преобразования $\mathbf{r}$ и t, кроме преобразования Галилея. На самом деле вторая формулировка *более общая*, чем предыдущая, так как в ней не конкретизирован вид того преобразования координат и времени, относительно которого инвариантны уравнения механики. В дальнейшем, если не сделано специальной оговорки, принцип относительности понимается во *второй формулировке.*

3. Принцип относительности отнюдь не утверждает, что одно и то же физическое явление выглядит одинаково в различных инерциальных системах отсчета. Дело в том, что одни только дифференциальные уравнения механики не определяют движение системы. К ним необходимо присоединить еще *начальные условия*, например задать координаты и скорости всех взаимодействующих частиц в определенный момент времени. А эти начальные условия меняются при переходе от одной системы отсчета к другой. Именно из-за различия начальных условий движение предмета, свалившегося с полки равномерно движущегося вагона, происходит вниз по прямой линии, если его рассматривать относительно самого вагона, тогда как относительно полотна железной дороги то же движение совершается по параболе. Вот почему в формулировке принципа относительности говорится не об одинаковости явлений, а об *одинаковости законов*, определяющих изменение состояний движения механических систем.

Смысл принципа относительности можно также уяснить на следующем примере. Возьмем замкнутую систему тел A и зададим их начальные положения и скорости относительно инерциальной системы отсчета S. Пусть имеется тождественная с A другая замкнутая система тел A', в которой созданы в точности такие же начальные условия, но уже относительно инерциальной системы отсчета S'. Тогда движение в системе тел A относительно S будет тождественным с движением в системе тел A' относительно S'. В этом и состоит *равноправие* инерциальных систем отсчета, устанавливаемое принципом относительности.

Принципиально взаимодействие имеет место между любыми телами Вселенной. Поэтому, если бы и существовали замкнутые системы, то единственной из таких систем могла бы быть, строго говоря, только вся Вселенная. Но тогда принцип относительности в том смысле, какой ему придан в последнем примере, был бы бессодержательным, так как двух Вселенных не существует. Принцип относительности только потому сохраняет содержание, что многими взаимодействиями, ввиду их слабости, можно пренебречь и таким путем выделить практически бесконечное множество ограниченных систем тел, каждая из которых приближенно ведет себя как *замкнутая.* В какой степени та или иная система удовлетворяет этому условию, зависит от исследуемого вопроса и от точности, какая предъявляется при его изучении. Достаточно ли для (приближенной) замкнутости системы отдаленности всех масс, не входящих в рассматриваемую систему? Ответ в соответствии с опытом гласит, что в случае *равномерного и прямолинейного движения* (относительно инерциальной системы отсчета) этого *достаточно*, а для всех остальных видов движения *недостаточно.*

4. Явления природы не представляется возможным разделить на чисто «механические» и «немеханические». Если бы даже это и можно было сделать, то принцип относительности не мог бы относиться к одним только «механическим» явлениям. Действительно, всякое «механическое» явление связано с множеством «физических» явлений и обусловлено ими. И если бы принцип относительности не был справедлив для этих «физических» явлений, то он не мог бы оставаться справедливым и для «чисто механических» явлений. Поэтому принцип относительности необходимо распространить на все явления природы и дать ему следующую формулировку:

Законы природы, определяющие изменение состояний физических систем, не зависят от того, к какой из двух инерциальных систем отсчета, движущихся одна относительно другой прямолинейно и равномерно, они относятся.

Это положение называется *частным*, или *специальным*, *принципом относительности Эйнштейна.* Он устанавливает равноправие *только инерциальных систем отсчета.* На основе этого принципа Эйнштейн создал в 1905 г. *частную*, или *специальную*, *теорию относительности.* Через 10 лет он обобщил принцип относительности на случай *произвольных неинерциальных систем отсчета* и создал *общую теорию относительности*, иначе называемую *релятивистской теорией тяготения.* Эта фундаментальная теория приобрела особое значение в связи с астрофизическими открытиями последнего времени. Общая теория относительности стала *основной теорией* в астрофизике, в частности в космологии. Однако в нашем курсе мы можем ограничиться изложением только специальной теории относительности.

§ 102. Опыт Майкельсона

1. Основные уравнения электродинамики Максвелла–Лорентца *не инвариантны относительно преобразования Галилея.* Действительно, скорость света в вакууме, вычисленная из этих уравнений, равна постоянной c. Такой результат оставался бы верным во всех инерциальных системах отсчета, если бы в них уравнения Максвелла имели один и тот же вид. Но это несовместимо с законом сложения скоростей (101.3), который является следствием преобразования Галилея.

Неинвариантность уравнений Максвелла относительно преобразования Галилея, как думали физики конца XIX и начала XX веков, не противоречит, однако, принципу относительности. Дело в том, что во всех электродинамических явлениях, помимо обыкновенного вещества, по господствовавшим в то время воззрениям, участвует еще *мировой эфир.* Уравнения Максвелла принимают обычный вид только в системе отсчета, в которой эфир покоится. Такая система является преимущественной или выделенной среди всех прочих инерциальных систем отсчета. Во всех остальных системах должно наблюдаться движение эфира — *эфирный ветер.* Его обнаружение считалось в XIX и начале XX веков одной из центральных проблем физики. Опыт привел к результату, что гипотеза эфира, как она понималась в физике XIX века, несостоятельна, а эфирного ветра не существует. Однако опыты по его обнаружению не совсем потеряли свое значение. Не вводя представления об эфире, можно поставить вопрос о существовании *преимущественной инерциальной системы отсчета.* На этот вопрос и отвечают опыты по обнаружению эфирного ветра.

2. Если принять гипотезу о существовании преимущественной системы отсчета, то прежде всего возникает вопрос, как движется Земля относительно этой системы? Обозначим скорость такого движения через $\mathbf{V}$. Тогда, согласно дорелятивистской кинематике, в направлении вектора $\mathbf{V}$ скорость света относительно Земли будет $c - V$, а в прямо противоположном направлении — $c + V$, где c — скорость света относительно преимущественной системы отсчета.

Этот результат можно подвергнуть экспериментальной проверке. Для этого, казалось бы, можно воспользоваться двумя одинаково идущими часами, установленными на Земле в различных точках A и B на расстоянии $l = AB$ друг от друга (рис. 323). Если прямая AB ориентирована вдоль скорости $\mathbf{V}$, то время, затрачиваемое светом на прохождение расстояния AB, будет $t_{AB} = l/(c-V)$, а на возвращение обратно $t_{BA} = l/(c+V)$. Разность этих времен, если ограничиться точностью до первого порядка относительно V/c, равна $t_{AB} - t_{BA} \approx 2\frac{l}{c}\frac{V}{c}$. Если прямая AB составляет с направлением скорости $\mathbf{V}$ угол α, то

$$t_{AB} - t_{BA} \approx 2\frac{l}{c}\frac{V}{c}\cos\alpha.$$

Время t_{AB} находится по разности показаний часов в моменты прихода светового сигнала в B и его отправления из A. Таким же путем находится время t_{BA}.

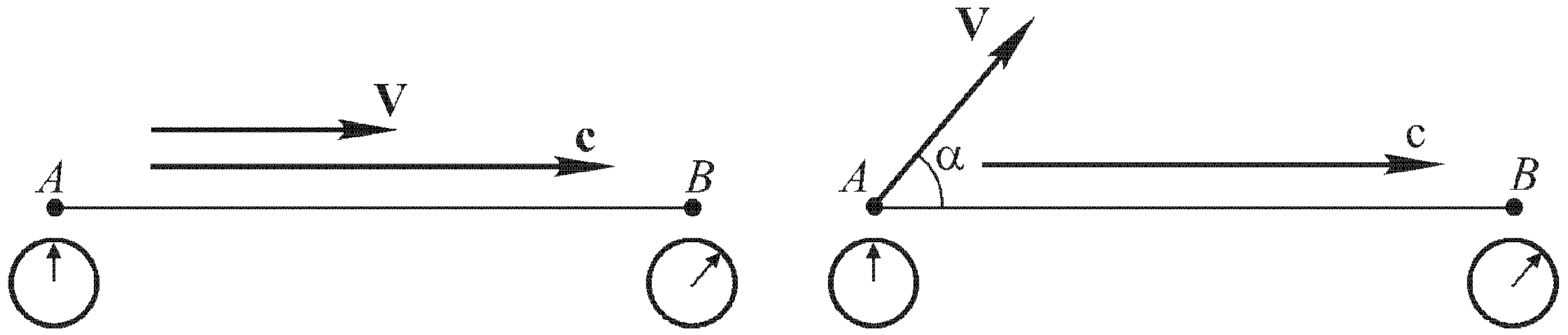


Рис. 323

Не было недостатка в предложениях, часто остроумных, как фактически осуществить опыт. Но в принципиальном отношении все они не отличались от схемы опыта, описанной выше. Предложения такого рода были оставлены после следующего замечания Майкельсона. Времена t_{AB} и t_{BA} измеряются с помощью *двух часов*, находящихся в различных точках пространства. Для измерения необходимо эти часы *синхронизовать*, что практически невозможно сделать с требуемой точностью. (Эйнштейн позднее заметил, что эта трудность *принципиальная*, а не только практическая, см. § 104.) Для осуществления опыта надо обойтись без синхронизации, т.е. все измерения производить с помощью *только одних часов*. Для этого из точки A, где помещены часы, надо послать световой сигнал и отразить его обратно в A с помощью зеркала, помещенного в B. Полное время распространения света t вдоль прямой AB туда и обратно можно измерить с помощью лишь одних часов. Делением удвоенного расстояния между точками A и B на это время можно найти скорость света c. Но это будет не скорость света в каком-либо одном определенном направлении, а *средняя скорость в двух прямо противоположных направлениях.* Майкельсон исследовал зависимость полного времени t от ориентации прямой AB.

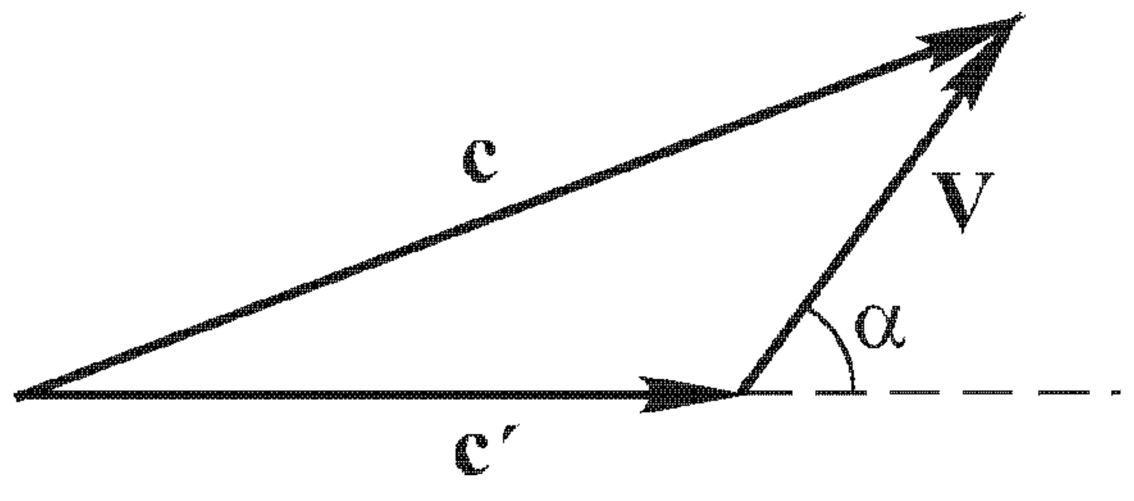


Рис. 324

Как видно из рис. 324, $c^2 = c'^2 + V^2 + 2Vc'\cos\alpha$, где $\mathbf{c}$ и $\mathbf{V}$ — скорости света и Земли относительно преимущественной системы отсчета, $\mathbf{c}'$ — скорость света относительно Земли, а α — угол между скоростями $\mathbf{V}$ и $\mathbf{c}'$. Решая полученное квадратное уравнение относительно c', найдем

$$c' = -V\cos\alpha + c\sqrt{1-\beta^2\sin^2\alpha},$$

где введено обозначение $\beta = V/c$. Ограничиваясь квадратами β, получим

$$c' = c\left(1 - \beta\cos\alpha - \frac{1}{2}\beta^2\sin^2\alpha\right).$$

3. Майкельсон воспользовался сконструированным им интерферометром, описанным в § 35 (см. рис. 140). Пластинка P_2 была введена для уравнивания оптических длин плеч интерферометра. Без этого интерференция в белом свете была бы невозможна. В принципиальных рассуждениях от наличия пластинки P_2 можно отвлечься и считать разделительную пластинку P_1 бесконечно тонкой. Пусть плечо интерферометра *2* составляет с направлением скорости $\mathbf{V}$ угол α. Тогда на прохождение расстояния от A до зеркала M_2 и обратно свет затратит время

$$t_2 = \frac{l}{c}\frac{1}{1-\beta\cos\alpha - {}^1\!/_2\beta^2\sin^2\alpha} + \frac{l}{c}\frac{1}{1+\beta\cos\alpha - {}^1\!/_2\beta^2\sin^2\alpha} =$$
$$= \frac{2l}{c}\frac{1-{}^1\!/_2\beta^2\sin^2\alpha}{(1-{}^1\!/_2\beta^2\sin^2\alpha)^2 - \beta^2\cos^2\alpha} \approx \frac{2l}{c}\frac{1-{}^1\!/_2\beta^2\sin^2\alpha}{1-\beta^2}.$$

Найдем теперь время t_1, которое требуется свету, чтобы пройти туда и обратно перпендикулярное плечо интерферометра. Оно найдется из предыдущего выражения заменой угла α на угол $\alpha \pm \frac{\pi}{2}$. Это дает

$$t_1 = \frac{2l}{c}\frac{1-{}^1\!/_2\beta^2\cos^2\alpha}{1-\beta^2}.$$

Таким образом,

$$t_2 - t_1 = \frac{l}{c}\frac{\beta^2}{1-\beta^2}\cos 2\alpha \approx \frac{l}{c}\beta^2\cos 2\alpha.$$

Этой разности времен соответствует разность фаз интерферирующих лучей

$$\varphi_2 - \varphi_1 = \frac{2\pi c}{\lambda}(t_2 - t_1) = \frac{2\pi l}{\lambda}\beta^2\cos 2\alpha.$$

При вращении интерферометра угол α будет меняться, а интерференционные полосы должны смещаться. Этот эффект и пытался обнаружить Майкельсон.

Если за единицу принять ширину интерференционной полосы, то смещение полос при повороте интерферометра будет

$$\Delta N = \frac{\Delta(\varphi_2 - \varphi_1)}{2\pi} = \frac{l}{\lambda}\beta^2\Delta(\cos 2\alpha). \qquad (102.1)$$

Максимального значения

$$\Delta N_{\text{макс}} = \frac{2l}{\lambda}\beta^2 \qquad (102.2)$$

оно достигает, когда при повороте интерферометра $\cos 2\alpha$ меняется от ± 1 до ∓ 1. Отсюда видно, что ожидаемое смещение очень мало, так как оно *второго порядка* по $\beta \equiv V/c$. Точно оценить его невозможно, поскольку скорость Земли относительно преимущественной инерциальной системы отсчета неизвестна. Можно, однако, оценить нижний предел $\Delta N_{\text{макс}}$. Действительно, допустим сначала, что Солнце покоится относительно преимущественной системы отсчета. Тогда под V надо понимать скорость движения Земли по своей орбите, которая равна около 30 км/с, так что $\beta = 10^{-4}$, $\beta^2 = 10^{-8}$. В первом опыте Майкельсона (1881 г.) длина плеча интерферометра была $l = 1{,}2$ м. Если взять $\lambda = 600$ нм, то получится $\Delta N_{\text{макс}} = 0{,}04$ полосы.

При учете движения Солнца можно ожидать и большего смещения. Действительно, разложим скорость движения Солнца на составляющую $\mathbf{V}_{\|}$ в плоскости земной орбиты и составляющую $\mathbf{V}_{\perp}$, перпендикулярную к этой плоскости. Полная скорость Земли будет $\mathbf{V} = \mathbf{V}_{\text{орб}} + \mathbf{V}_{\|} + \mathbf{V}_{\perp}$, где $\mathbf{V}_{\text{орб}}$ — скорость орбитального движения Земли. Если опыт производится в момент, когда направления скоростей $\mathbf{V}_{\text{орб}}$ и $\mathbf{V}_{\|}$ совпадают между собой, то для полной скорости Земли можно написать

$$V^2 = (V_{\text{орб}} + V_{\|})^2 + V_{\perp}^2,$$

так что V получится больше $V_{\text{орб}}$. Следовательно, и ожидаемое смещение полос будет больше вычисленного выше.

Конечно, минимальное ожидаемое смещение в 0,04 полосы, вычисленное выше, могло бы наблюдаться только два раза в год, когда направление эфирного ветра совпадает или прямо противоположно направлению движения Земли. Поэтому необходимо производить наблюдения в течение по крайней мере *полугода*, что фактически и делалось Майкельсоном и всеми физиками, повторявшими его опыт.

Опыт был поставлен настолько тщательно, что позволял обнаружить смещение на 0,01 интерференционной полосы. Ожидаемое смещение на 0,04 полосы могло быть уверенно обнаружено на фоне неизбежных случайных ошибок измерений. Оказалось, что смещения нет. Точнее, результат опыта позволял утверждать, что скорость эфирного ветра, если бы он существовал, не может превышать 18 км/с.

4. В 1887 г. Майкельсон повторил свой опыт совместно с Морли (1837–1925). Для того чтобы при вращении интерферометра не возникали деформации, вызывавшие небольшое смещение интерференционных полос, интерферометр вместе с остальной аппаратурой монтировался на тяжелой цементной плите, которая плавала в сосуде со ртутью. С той же целью прибор все время находился в медленном вращении, даже в моменты наблюдения. Путем применения многократных отражений от 16 зеркал, установленных на той же плите, эффективная длина плеча интерферометра была доведена до 11 м. Ожидаемое смещение из-за орбитального движения Земли составляло уже 0,4 интерференционной полосы. Максимально наблюдавшееся смещение (из-за неизбежных случайных помех) составляло 0,02, а среднее — много меньше 0,01 интерференционной полосы. Опыт позволил утверждать, что скорость эфирного ветра не сможет превышать 7 км/с. Сам Майкельсон пришел к заключению, что результат опыта определенно отрицательный — *эфирного ветра не существует.*

Опыт Майкельсона неоднократно повторялся другими физиками с улучшенной аппаратурой и улучшенными методами наблюдения. Не останавливаясь на истории вопроса, укажем только, что к началу 30-х годов было с достоверностью установлено, что скорость эфирного ветра, если он и существует, не может превышать 1–1,5 км/с.

5. Изобретение лазеров, благодаря исключительно высокой степени монохроматичности и пространственной когерентности этих источников света, позволило чрезвычайно упростить схему опыта Майкельсона и повысить его точность. Опыт был поставлен в 1964 г. американскими физиками Джасея, Джаваном, Мурреем и Таунсом. Идея опыта заключалась в следующем.

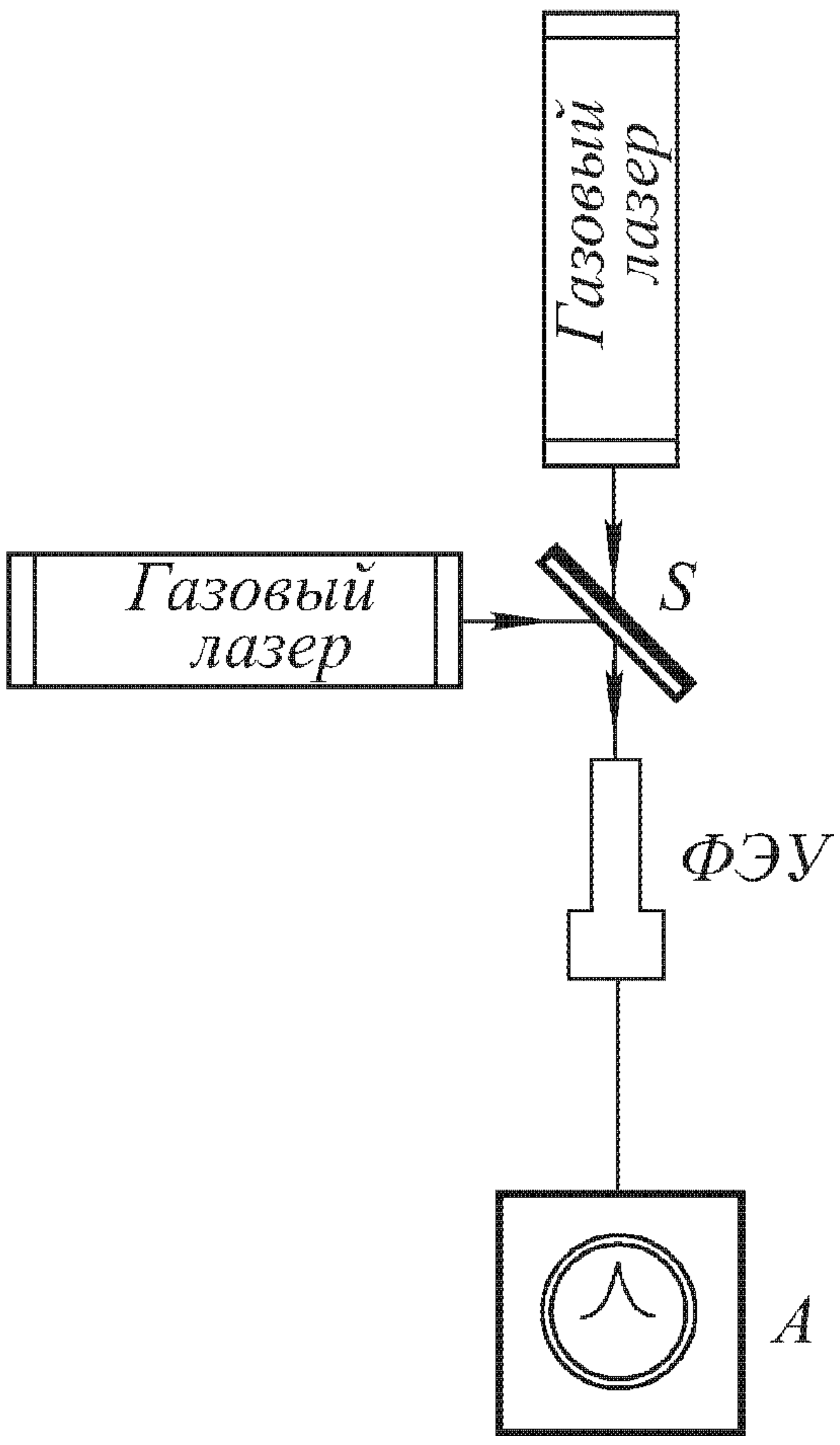


Рис. 325

Если скорость света зависит от направления его распространения относительно Земли, то частота, на которой генерирует лазер, должна изменяться при его повороте. Это изменение пропорционально β^2, т.е. второго порядка по β. Два одинаковых гелий-неоновых лазера были установлены на поворотной платформе перпендикулярно друг к другу (рис. 325). Световой пучок от одного из лазеров проходил полупосеребренное зеркало S, а пучок от другого лазера отражался от того же зеркала. Далее пучки шли в одном и том же направлении и попадали в фотоэлектронный умножитель, установленный на той же поворотной платформе. Если бы частоты этих пучков немного отличались друг от друга, то должны были бы возникнуть биения фототока с частотой, лежащей в области радиодиапазона, которые можно было бы наблюдать обычными радиотехническими приемами с помощью анализатора A. Частота биений должна была бы меняться при вращении прибора. Если в исходном положении один лазер был ориентирован вдоль, а другой — перпендикулярно к направлению движения Земли, то при повороте прибора на 90°, согласно теории неподвижного эфира, из-за орбитального движения Земли должно было бы наблюдаться изменение разности частот лазеров примерно на 3 МГц, тогда как возможная ошибка опыта не превосходила нескольких герц. На опыте такое смещение обнаружено не было. На основании своих наблюдений экспериментаторы пришли к заключению, что скорость эфирного ветра, если бы таковой существовал, не может превышать 30 м/с. По сравнению с предыдущими результатами точность была повышена примерно в 50 раз.

6. Отрицательный результат опыта Майкельсона доказывает, что в земной системе отсчета скорость света в вакууме *изотропна*, т.е. не зависит от направления его распространения. Под скоростью света здесь понимается *средняя скорость* в двух прямо противоположных направлениях — она и только она определяет результат опыта с «одними часами». Обобщая этот результат, можно сказать, что он справедлив *во всех инерциальных системах отсчета.*

Остается выяснить, одинаковы ли числовые значения скорости света в различных системах отсчета? Утвердительный ответ можно дать на основании опыта Кеннеди и Торндайка, которые в 1932 г. изменили опыт Майкельсона. В их опыте использовался интерферометр Майкельсона, в котором длины плеч были *разными* и отличались на 16 см. Они наблюдали интерференционные полосы (при неподвижном относительно Земли интерферометре) в течение шести месяцев. За это время Земля из своего исходного положения непрерывно перемещалась в диаметрально противоположную точку своей орбиты, участвуя вместе с тем в движении Солнца. Тем самым реализовалось множество систем отсчета, отличающихся друг от друга положением и ориентацией в пространстве. Если бы в них скорость света была разной, то должно было бы наблюдаться регулярное смещение интерференционных полос. На самом деле смещение носило случайный характер и не превосходило 3/1000 ширины полосы. На этом основании можно прийти к заключению, что с точностью примерно 2 м/с скорость света в вакууме относительно Земли одинакова во всех ее положениях в пространстве.

Все это подтверждает равноправие различных инерциальных систем отсчета по отношению ко всем физическим явлениям — *преимущественной инерциальной системы отсчета не существует.*

§ 103. Независимость скорости света от движения источника

1. Помимо принципа или постулата относительности, Эйнштейн положил в основу теории относительности *постулат о независимости скорости света в вакууме от движения источника.* Его обычно называют *принципом «постоянства скорости света»*, хотя, как заметил еще Паули (1900–1958), такое название может привести к недоразумениям. Об универсальности постоянства скорости света в вакууме не может быть речи уже потому, что эта скорость постоянна *только в инерциальных системах отсчета*, а независимость скорости света от движения источника сохраняется и *в общей теории относительности*, где на выбор систем отсчета не накладывается никаких ограничений.

Независимость скорости света от движения источника можно было бы не выдвигать в качестве самостоятельного постулата, если с самого начала принять волновую теорию света. Однако столь фундаментальную теорию, какой является теория относительности, лучше строить, не связывая ее ни с какими гипотетическими представлениями о природе и механизме физических явлений.

2. Предположение о зависимости скорости света от движения источника ввел Ритц (1878–1909). Согласно этому предположению скорость света в вакууме постоянна и равна c только относительно источника. Если же источник света движется со скоростью $\mathbf{v}$ относительно какой-либо системы отсчета, то скорость света $\mathbf{c}'$ в этой системе векторно складывается из $\mathbf{c}$ и $\mathbf{v}$, т.е. $\mathbf{c}' = \mathbf{c} + \mathbf{v}$, как это происходит со скоростью снаряда при стрельбе из движущегося орудия. Такая гипотеза получила название *баллистической.* Ритц перестроил электродинамику Максвелла, приведя ее в соответствие с этой гипотезой. Ясно, что баллистическая гипотеза объясняет отрицательный результат опыта Майкельсона, если только источник света неподвижен относительно интерферометра.

Но баллистическая гипотеза, как и любое предположение о зависимости скорости света от движения источника, не согласуется с молекулярным объяснением отражения и преломления света (см. § 68 и 69). В основе такого объяснения лежит представление, что отраженная и преломленная волны возникают в результате интерференции падающей волны с вторичными волнами, излучаемыми молекулами и атомами среды. Действительно, если среду считать неподвижной, а источник света взять движущимся, то волны, излучаемые источником, будут распространяться с иными скоростями, чем волны, излучаемые молекулами н атомами среды. Интерференция между ними невозможна. Однако это замечание есть аргументация, основанная на волновой теории света, а не экспериментальное доказательство, свободное от гипотетических представлений о природе света.

Убедительное доказательство несостоятельности баллистической гипотезы, как показал в 1913 г. голландский астроном де Ситтер (1872–1934), дают астрономические наблюдения над движением двойных звезд. Действительно, допустим, что баллистическая гипотеза верна. Предположим для простоты, что компоненты двойной звезды вращаются вокруг их центра масс по круговым орбитам в той же плоскости, в которой расположена Земля. Рассмотрим движение одной из этих двух звезд. Пусть v — скорость движения ее по круговой орбите. В положении звезды, когда она удаляется от Земли вдоль соединяющей их прямой, скорость света равна $(c - v)$, а в положении, когда звезда приближается, равна $(c + v)$. Если отсчитывать время от момента, когда звезда находилась в первом положении, то свет из этого положения дойдет до Земли в момент $t_1 = L/(c - v)$, а из второго положения — в момент $t_2 = T/2 + L/(c + v)$, где T — период обращения звезды, а L — расстояние до нее. При громадных расстояниях до звезд наблюдаемые движения звезды могли бы заметно отступать от законов Кеплера. В частности, при очень больших L могло бы случиться, что $t_2 \leqslant t_1$, т.е. звезда одновременно была бы видна в двух (и даже нескольких) положениях или обращалась бы в противоположном направлении. Ничего подобного, как показали астрономические наблюдения, не происходит.

Если бы баллистическая гипотеза была верна, то опыт Майкельсона должен был бы дать положительный результат, если его произвести не с земным источником света, а со светом от звезд. Томашек в 1926 г. так и поступил, но получил отрицательный результат.

А.М. Бонч-Бруевич и В.А. Молчанов в 1956 г., применив современные методы измерения скорости света, сравнили скорости света от правого и левого краев Солнца. Из-за осевого вращения Солнца один из этих краев приближался к нам со скоростью 2,3 км/с, а другой удалялся с той же скоростью. Обе скорости света с достаточной точностью совпали между собой.

Наконец, ставились специальные опыты, в которых сравнивались скорости γ-квантов, испускаемых движущимися возбужденными ядрами углерода ($^*\mathrm{C}^{12}$) и неподвижными возбужденными ядрами кислорода ($^*\mathrm{O}^{16}$). Ставились также опыты по аннигиляции электрона с позитроном, специально приспособленные для проверки независимости скорости света от движения источника. По сравнению с астрономическими опытами, в которых используются космические источники света со сравнительно малыми скоростями, в опытах с атомными ядрами и элементарными частицами скорость источников гораздо выше (сравнима со скоростью самого света). Опыты подтвердили, что с точностью около 10 % скорость γ-квантов не зависит от движения источников.

Все приведенные опыты подтверждают постулат о независимости скорости света в вакууме от движения источника.

3. Заметим в заключение, что теория относительности вообще была бы невозможна, если бы не был установлен фундаментальный факт *конечности скорости распространения света*. Изучение методов и результатов измерения скорости света представляет громадный, не только исторический интерес. В частности, уточнение числового значения этой постоянной необходимо для точных измерений астрономических расстояний методами радиолокации. Это в свою очередь необходимо для целей космонавтики. Однако мы не можем касаться этих вопросов. Ограничимся замечанием, что в 1972 г. скорость света была определена на основе независимых измерений длины волны λ и частоты света ν. Источником света служил гелий-неоновый лазер, генерировавший излучение с длиной волны 3,39 мкм. Длина волны измерялась интерферометрически сравнением ее с эталоном длины, т.е. с длиной волны в вакууме оранжевой линии изотопа криптона-86. Ошибка таких измерений $\sim 10^{-5}$ нм. Частота лазерного излучения измерялась путем сравнения ее с атомным стандартом частоты, т.е. с частотой перехода между двумя сверхтонкими квантовыми уровнями атома цезия-133 в нулевом магнитном поле. При этом использовались методы нелинейной оптики — генерация излучений с суммарной и разностной частотами. В итоге для скорости света $c = \lambda\nu$ было найдено значение

$$c = 299\,792\,458 \pm 1{,}2 \text{ м/с},$$

превосходящее по точности все ранее полученные значения более чем на два порядка.

§ 104. Понятие одновременности

1. До теории относительности считалось, что принцип относительности и принцип независимости скорости света от движения источника противоречат друг другу. Действительно, согласно принципу относительности скорость света в вакууме во всех инерциальных системах отсчета должна быть одной и той же. Обозначим ее, как всегда, через c. Возьмем две инерциальные системы отсчета S и S', движущиеся относительно друг друга прямолинейно и равномерно. Пусть O и O' — какие-либо две точки, неподвижные в системах S и S' соответственно и пространственно совпадающие в момент времени $t = 0$. Произведем в этих точках в тот же момент времени кратковременную световую вспышку. Рассмотрим распространение этой вспышки сначала в системе S. Так как свет во всех направлениях распространяется с одной и той же скоростью c, то к моменту времени t вспышка достигнет сферы Σ радиуса ct с центром в точке O. Рассмотрим теперь распространение той же вспышки в системе S'. Обе системы совершенно равноправны. Они отличаются друг от друга только тем, что источник света относительно них движется по-разному. Но это в рассматриваемом вопросе не имеет никакого значения, так как скорость света не зависит от движения источника. Поэтому можно утверждать, что к моменту t та же вспышка достигнет сферы Σ' с тем же радиусом ct, но с центром в точке O'. Сферы Σ и Σ' разные, так как к моменту t их центры O и O' разойдутся. Таким образом, одна и та же световая вспышка *одновременно* должна находиться и на сфере Σ, и на сфере Σ', что, очевидно, невозможно.

2. Однако приведенное рассуждение несостоятельно, так как в его основе лежит *интуитивное представление об одновременности.* Если два события происходят в одном и том же месте пространства, то не возникает трудностей при суждении о том, одновременны эти события или нет. Но если события происходят *в разных местах пространства*, то без дальнейших разъяснений утверждение об их одновременности или неодновременности просто бессодержательно, так как *никакой априорной одновременности пространственно разделенных событий не существует.*

В принципе пространственно-временное описание явлений природы можно дать и без использования понятия одновременности. Действительно, возьмем какую-либо инерциальную систему отсчета. Для определения положения точки в ней будем пользоваться координатными осями, реализованными, например, с помощью трех взаимно перпендикулярных твердых стержней. Расстояние между двумя точками будем измерять путем откладывания неподвижного масштабного стержня, т.е. твердого стержня определенной длины, разделенного на метры и более мелкие единицы. Для измерения больших расстояний этот способ непригоден. Тогда надо применять косвенные методы измерения длин, о которых говорилось в § 1 первого тома. Для определения времени расставим в нашей координатной системе достаточно часто какие-либо *одинаково устроенные часы.* Стрелки часов в различных точках пространства установим совершенно произвольно и независимо друг от друга. После этого получится пространственно-временная система отсчета. Всякое событие в такой системе можно характеризовать местом, где оно произошло, и показаниями часов в том же месте.

Пока что введено только *местное время*, т.е. время в каждой точке пространства, причем времена в различных точках никак не связаны между собой. Это обстоятельство делает нашу систему практически совершенно непригодной для пространственно-временного описания явлений природы. Практической может быть только система с *единым временем*, в которой показания часов в различных точках не независимы, а определенным образом связаны или синхронизированы друг с другом. Способ синхронизации в принципе произволен. Требуется только, чтобы он был *внутренне непротиворечив*. Однако соображения целесообразности, основанные на принципе относительности и независимости скорости света от движения источника, почти однозначно вынуждают принять способ синхронизации, предложенный Эйнштейном, и основанное на нем определение одновременности.

3. Эйнштейн предложил устанавливать синхронизацию пространственно разделенных часов с помощью световых сигналов в вакууме. Пока синхронизация часов не установлена, не имеет смысла говорить о скорости в любом определенном направлении. Имеет смысл говорить лишь о *средней скорости* света в двух прямо противоположных направлениях, так как эту величину можно измерить с помощью *только одних часов*. Для этого из неподвижной точки A в момент времени t_1^A пошлем световой сигнал к неподвижному зеркалу, установленному на расстоянии l от A. Пусть отраженный сигнал вернется в A в момент t_2^A. Время отсчитывается по одним и тем же часам, так что никакой синхронизации не требуется. Разделив пройденное расстояние $2l$ на затраченное время $t_2^A - t_1^A$, найдем искомую среднюю скорость света $c = 2l/(t_2^A - t_1^A)$. Опыт Майкельсона и другие опыты показали, что эта величина есть *универсальная постоянная*. Теперь *полагаем по определению, что скорости света во всяких двух взаимно противоположных направлениях одинаковы*. Тем самым постулативно будет установлена *синхронизация часов* и введено *понятие одновременности*. После этого уже имеет смысл говорить о скорости любого движения в *одном направлении*, так как время, затрачиваемое на такое движение, можно измерить по двум пространственно разделенным часам, синхронизованным между собой.

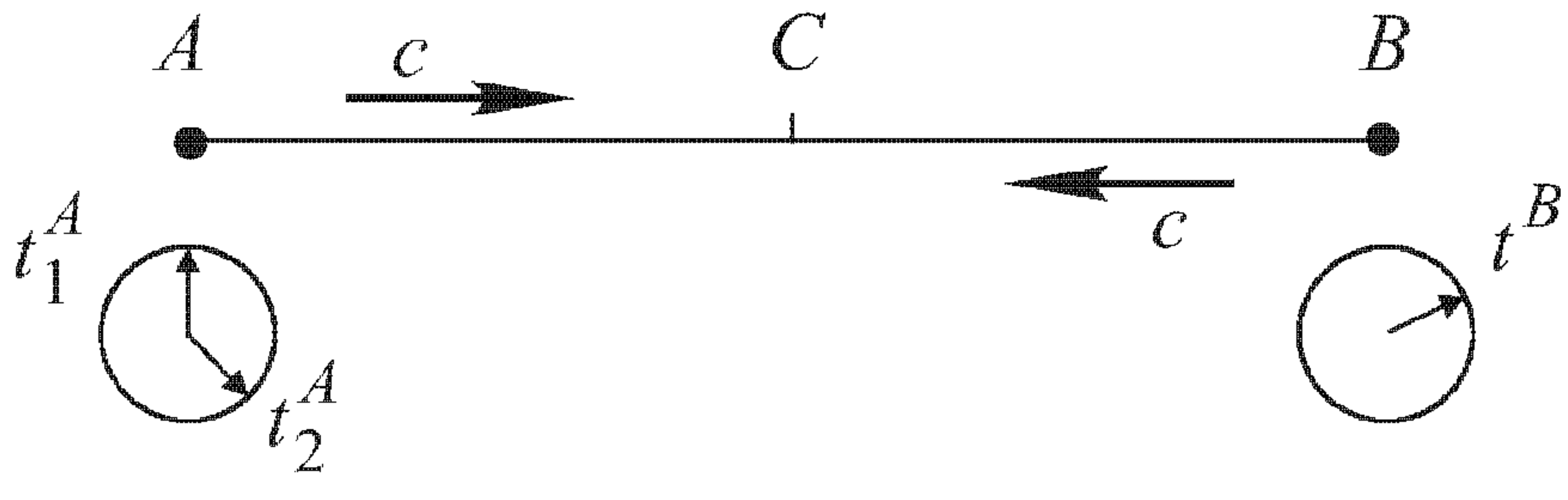


Рис. 326

Сформулируем теперь окончательно *правило Эйнштейна синхронизации часов. Пусть в неподвижных точках A и B* (рис. 326) *установлены одинаковые часы. Из точки A в момент t_1^A по часам A пошлем световой сиг-*

нал к зеркалу, установленному в точке B. Пусть отраженный сигнал возвращается в A в момент t_2^A (по часам A). В момент прихода сигнала в B поставим на часах в B время $t^B = {}^1\!/_2(t_1^A + t_2^A)$. Тогда по определению часы в A и B будут синхронизованы между собой.

Можно дать другой, эквивалентный, способ синхронизации. *В середине отрезка AB произведем световую вспышку. В момент прихода света от вспышки к часам A и B поставим их одинаково. Тогда часы будут синхронизованы.*

Но часы A и B можно синхронизовать между собой не только непосредственно, но также посредством третьих часов C. Синхронизуем сначала одним из указанных способов часы A и C, затем часы B и C. Мы утверждаем и рассматриваем это утверждение как постулат, что при этом часы A и B окажутся синхронизованными. Такое утверждение, как показывают все следствия из него, не приводит к противоречиям и согласуется с опытом. Таким образом, получается *пространственно-временная система отсчета с единым временем*, в которой часы синхронизованы между собой по правилу Эйнштейна. *Два пространственно разделенных события в этой системе отсчета называются одновременными, если часы, находящиеся в тех точках, где происходят эти события, показывают одно и то же время.*

4. *Одновременность есть понятие относительное*, т.е. два события, одновременные в одной инерциальной системе отсчета, могут оказаться не одновременными в другой системе. Поясним это на следующем мысленном опыте. По поверхности Земли равномерно и прямолинейно со скоростью V движется поезд $A'B'$ (рис. 327). Примем его за движущуюся систему отсчета S'. Неподвижной системой S будем считать полотно железной дороги. Пусть в концы поезда A' и B' ударяют две молнии, оставляющие следы A и B на полотне железной дороги. Отметим в поезде среднюю точку C', а на полотне — среднюю точку C между A и B. Пусть в точку C вспышки от молний приходят одновременно. Тогда с точки зрения системы S удары молний в концы поезда будут событиями одновременными. Но в момент встречи обеих вспышек точка C' окажется правее C. В этот момент вспышка от B' уже прошла через C', а от A' еще не дошла. Значит, с точки зрения системы S' удар молнии в B' произошел раньше удара в A'.

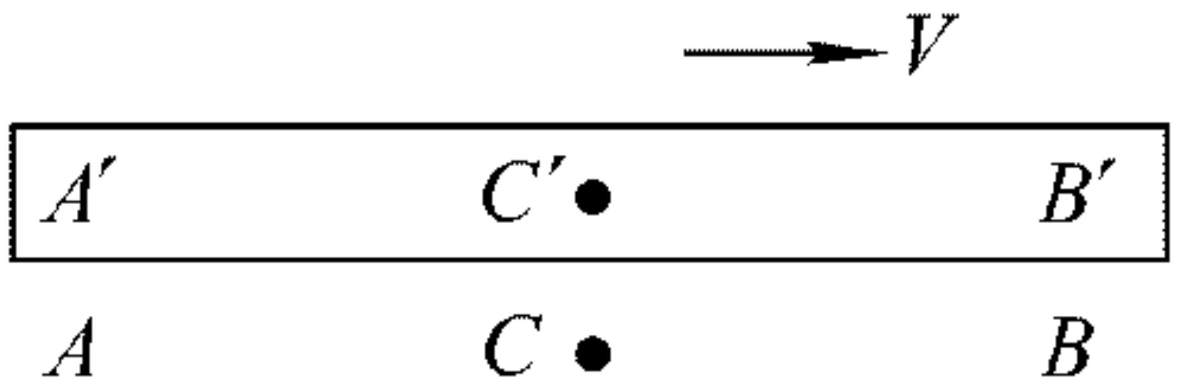


Рис. 327

Теперь ясно, как разрешается парадокс, о котором говорилось в п. 1. Свет доходит до сферы Σ одновременно в системе отсчета S, но не одновременно в системе S'. Аналогично, до сферы Σ' свет доходит одновременно с точки зрения системы отсчета S', но не одновременно с точки зрения системы S.

Дорелятивистская физика развивалась, и вполне успешно, считая время и одновременность *абсолютными*, т.е. одинаковыми во всех системах отсчета. Но так происходило лишь до тех пор, пока рассматривались *медленные движения*. А распространение света есть быстрый процесс. Вот почему именно в оптике физика встретилась раньше всего с принципиальными трудностями, преодоленными теорией относительности.

§ 105. Преобразование координат и времени в теории относительности

1. Рассмотрим две инерциальные системы отсчета S и S', из которых вторая движется относительно первой прямолинейно и равномерно со скоростью $\mathbf{V}$, а следовательно, первая движется относительно второй со скоростью $-\mathbf{V}$. В каждой системе отсчета расставлены достаточно часто одинаковые часы, неподвижные в этой системе и синхронизованные по правилу Эйнштейна. Пусть x, y, z, t — координаты и время какого-либо события (например, столкновения двух шаров) в системе отсчета S, а x', y', z', t' — координаты и время того же события в системе отсчета S'. Возникает вопрос, как по значениям x, y, z, t найти значения x', y', z', t' и наоборот. Решение этого вопроса основано на предположении, что пространство *однородно* и *изотропно*, а время *однородно*[1]. Однородность пространства и времени означает, что все точки пространства и все моменты времени, как в системе S, так и в системе S', абсолютно эквивалентны. Изотропия же пространства означает полную эквивалентность всех пространственных направлений в системе S, а также в системе S'. В силу указанной однородности и изотропии пространства и времени связь между x, y, z, t и x', y', z', t' должна быть *линейной*.

Вопрос о чисто пространственных преобразованиях координат в этих предположениях решается в аналитической геометрии. Поэтому можно отвлечься от этого вопроса и сосредоточить все внимание на том, что нового вносит в преобразование координат и времени равномерное движение одной системы отсчета относительно другой. Для этой цели достаточно рассмотреть частный случай, когда начала O и O' координатных систем S

---

[1] Эти свойства нельзя рассматривать как априорные. Они установлены и проверены экспериментально. Их объяснение, по-видимому, должна дать космология при рассмотрении происхождения и эволюции Вселенной в целом.

и S' в некоторый момент времени совмещаются. Этот момент мы примем за начало отсчета времени как в системе S, так и в системе S'. Тогда связь между x, y, z, t и x', y', z', t' будет не только линейной, но и *однородной*, так как нулевым значениям нештрихованных параметров соответствуют также нулевые значения штрихованных. Кроме того, оси X', Y', Z' координатной системы S' можно выбрать так, чтобы они были параллельны осям X, Y, Z координатной системы S и, следовательно, ось X' все время совмещалась с осью X (рис. 328).

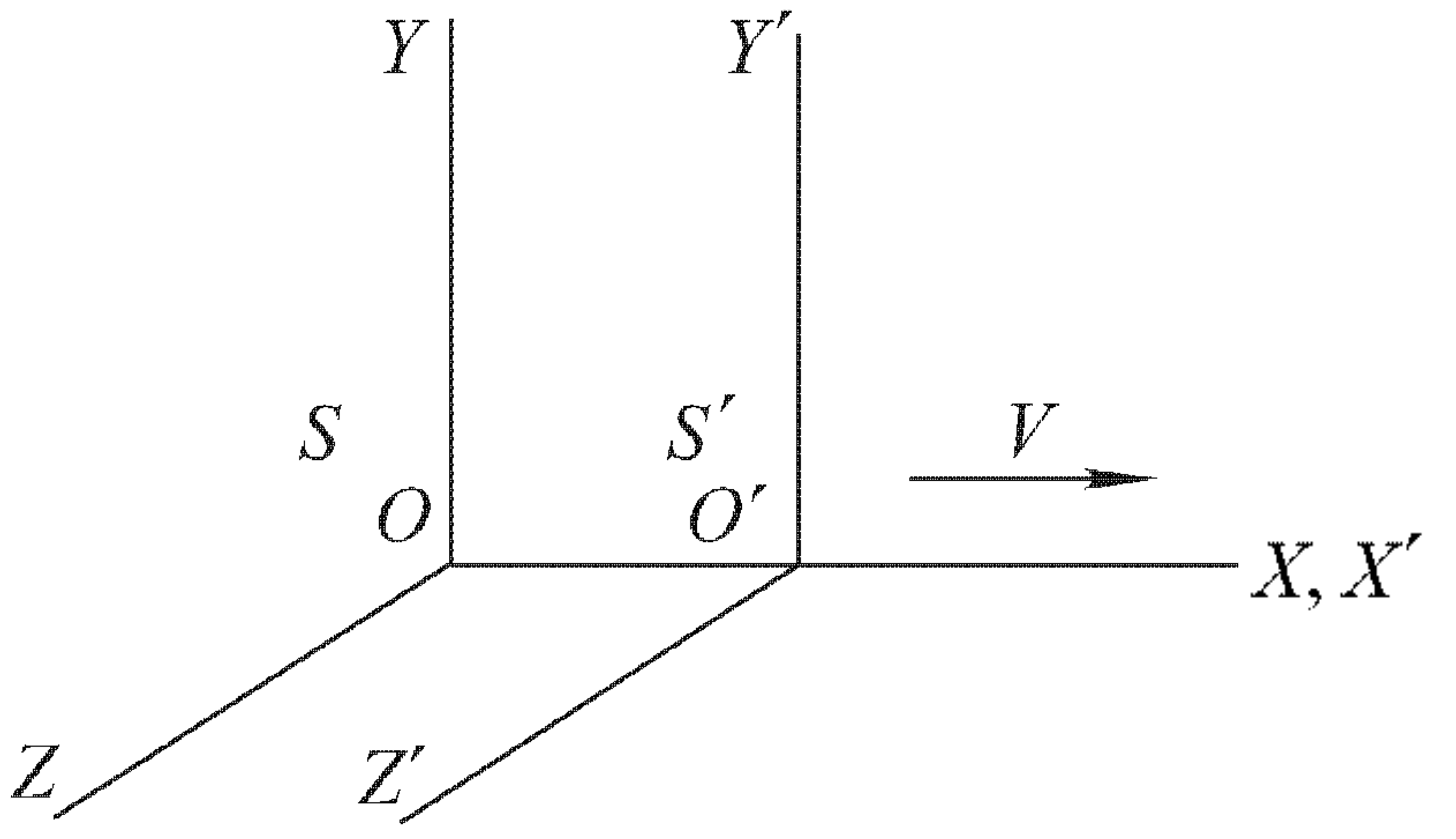


Рис. 328

2. Предположим, что относительная скорость V координатных систем S и S' меньше скорости света c $(V < c)$. Пусть при этом условии из начала координат O в момент времени $t_1 > 0$ (по часам системы S) послан световой сигнал в положительном направлении оси X. Пусть этот сигнал приходит в точку O' в момент времени t' (по часам системы S'). Тогда, ввиду линейности связи между координатами и временем в системах S и S', время t' должно выражаться также линейно через t_1. При этом нулевому значению t_1 соответствует нулевое значение t', так как в момент $t = t'$ начала координат O и O' совмещаются между собой. Следовательно, должно быть

$$t' = kt_1, \tag{105.1}$$

где k — некоторый коэффициент. В силу изотропии пространства он может зависеть только от абсолютного значения скорости $\mathbf{V}$, но не от ее направления.

Выразим коэффициент k через скорость V. Пусть в точке O' световой сигнал отражается и, распространяясь в обратном направлении, возвращается в точку O в момент t_2 (по часам системы S). Тогда, ввиду полной эквивалентности систем отсчета S и S',

$$t_2 = kt', \tag{105.2}$$

где k имеет то же значение, что и в формуле (105.1). Исключив t', получим

$$t_2 = k^2 t_1. \tag{105.3}$$

Согласно эйнштейновскому правилу синхронизации часов сигнал, посланный из O в момент t_1 и возвратившийся обратно в момент t_2, отражается в точке O' в момент времени

$$t = \frac{t_1 + t_2}{2} = \frac{1 + k^2}{2} t_1 \tag{105.4}$$

(по часам системы S). Расстояние x, проходимое точкой O' за время t, равно $x = Vt$. То же расстояние свет проходит за время $t - t_1$, и следовательно, $x = c(t - t_1)$. Таким образом, $Vt = c(t - t_1)$. Подставив сюда значение t из предыдущей формулы, получим

$$V(k^2 + 1) = c(k^2 - 1), \tag{105.5}$$

откуда

$$\frac{V}{c} = \frac{k^2 - 1}{k^2 + 1}, \tag{105.6}$$

$$k^2 = \frac{c + V}{c - V} = \frac{1 + V/c}{1 - V/c}. \tag{105.7}$$

Извлекая из последнего выражения квадратный корень, найдем k. После этого получаем две вспомогательные формулы:

$$k + \frac{1}{k} = \frac{2}{\sqrt{1 - (V/c)^2}}, \quad k - \frac{1}{k} = \frac{2V/c}{\sqrt{1 - (V/c)^2}}, \tag{105.8}$$

которые понадобятся нам в дальнейшем. Все эти соотношения, разумеется, имеют смысл только при условии $V < c$.

Отметим, в частности, что из формул (105.1) и (105.4) следует

$$\frac{t'}{t} = \frac{2k}{k^2 + 1} = \sqrt{1 - (V/c)^2}, \tag{105.9}$$

откуда $t' < t$. Величина t есть время движения точки O' из неподвижного начала O до точки, в которой его догонит световой сигнал, измеренное по «неподвижным часам», т.е. часам системы S. Величина t' имеет смысл того же времени, но измеренного уже по «движущимся часам», т.е. часам системы S'. Таким образом, величины t и t' представляют собой времена между одними и теми же событиями, измеренные соответственно в «неподвижной» S и «движущейся» S' системах отсчета. Формула (105.9) показывает, что движущиеся часы идут медленнее неподвижных. Это явление будет обсуждено ниже с различных точек зрения, а сейчас мы вернемся к вопросу о преобразовании координат и времени.

3. Пусть на оси X произошло какое-то событие A, например столкновение двух шаров (рис. 329). В системе отсчета S это событие характеризуется абсциссой x и моментом времени t. В системе отсчета S' абсцисса и время того же события будут x', t'. Рассмотрим световой сигнал, отправляющийся из начала координат O к месту события A (рис. 329 *а*). На прохождение расстояния OA сигнал затрачивает время x/c. Чтобы он пришел

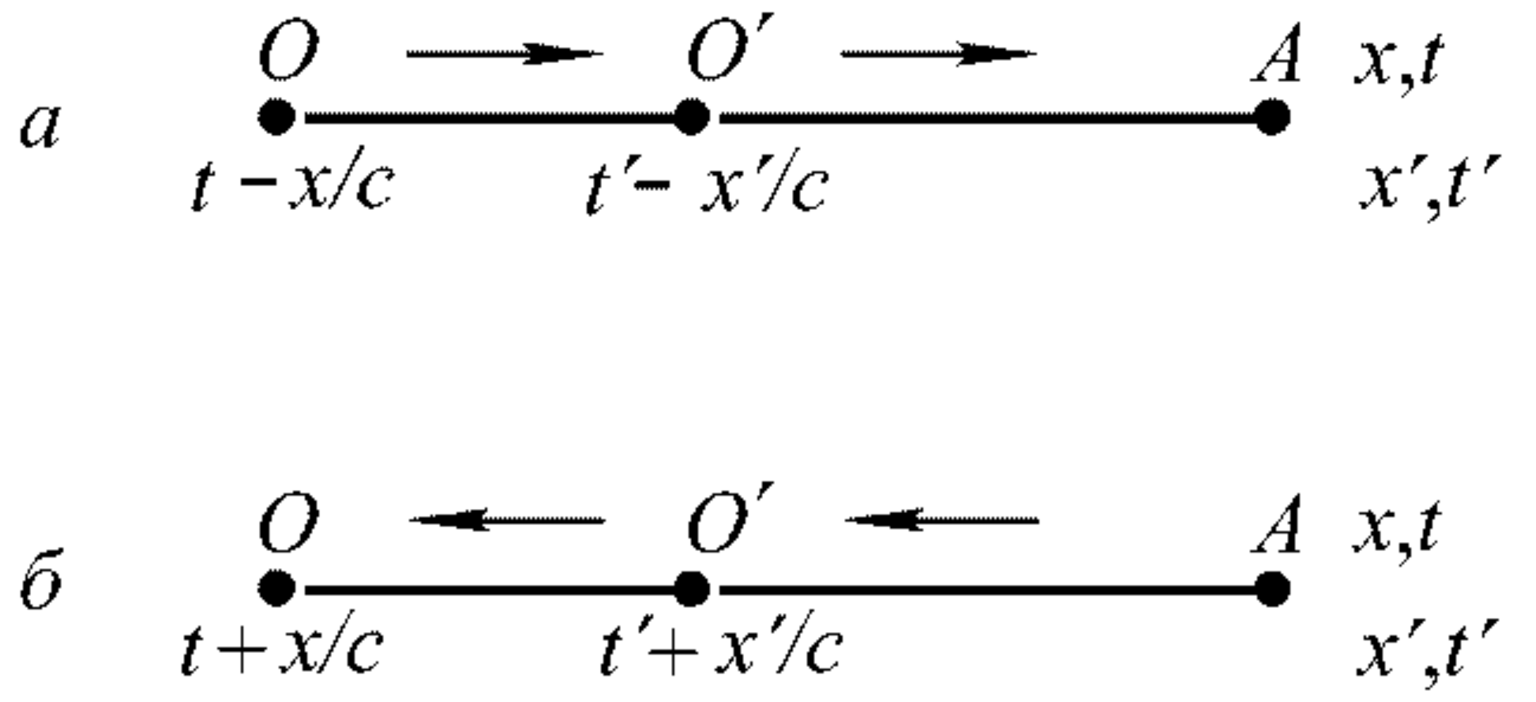


Рис. 329

в A в момент t, его надо отправить из O в момент $t_1 = t - x/c$. На прохождение расстояния $O'A$ тот же сигнал затрачивает время x'/c, так что он проходит через точку O' в момент $t_1' = t' - x'/c$. По определению коэффициента k $t_1' = kt_1$ или $t' - x'/c = k(t - x/c)$. Отразим теперь световой сигнал в A в обратном направлении (рис. 329 *б*). Через точку O' он пройдет в момент времени $t_2' = t' + x'/c$, а через точку O — в момент $t_2 = t + x/c$. Теперь роли штрихованных и нештрихованных величин поменялись местами, и на основании того же определения коэффициента k можно написать $t_2 = kt_2'$, или $t + x/c = k(t' + x'/c)$. Таким образом,

$$t' - x'/c = k(t - x/c), \qquad k(t' + x'/c) = t + x/c. \tag{105.10}$$

Полученные соотношения справедливы, в каком бы месте оси X ни произошло событие A. Пусть, например, оно произошло левее точек O и O', как указано на рис. 330 *а*. Отправим от места события световой сигнал вправо. Так как теперь x и x' отрицательны, то сигнал достигнет точек O и O' в более поздние моменты $t - x/c$ и $t' - x'/c$. На основании определения коэффициента k напишем $t' - x'/c = k(t - x/c)$, а это есть первое соотношение (105.10). Возьмем теперь световой сигнал, распространяющийся справа налево (рис. 330 *б*). Пусть он достигает точки A в моменты t и t' по часам в системах S и S' соответственно. Через точки O и O' сигнал пройдет в моменты времени $t + x/c$ и $t' + x'/c$, а потому $t' + x'/c = k(t' - x'/c)$, т.е. получается и второе соотношение (105.10).

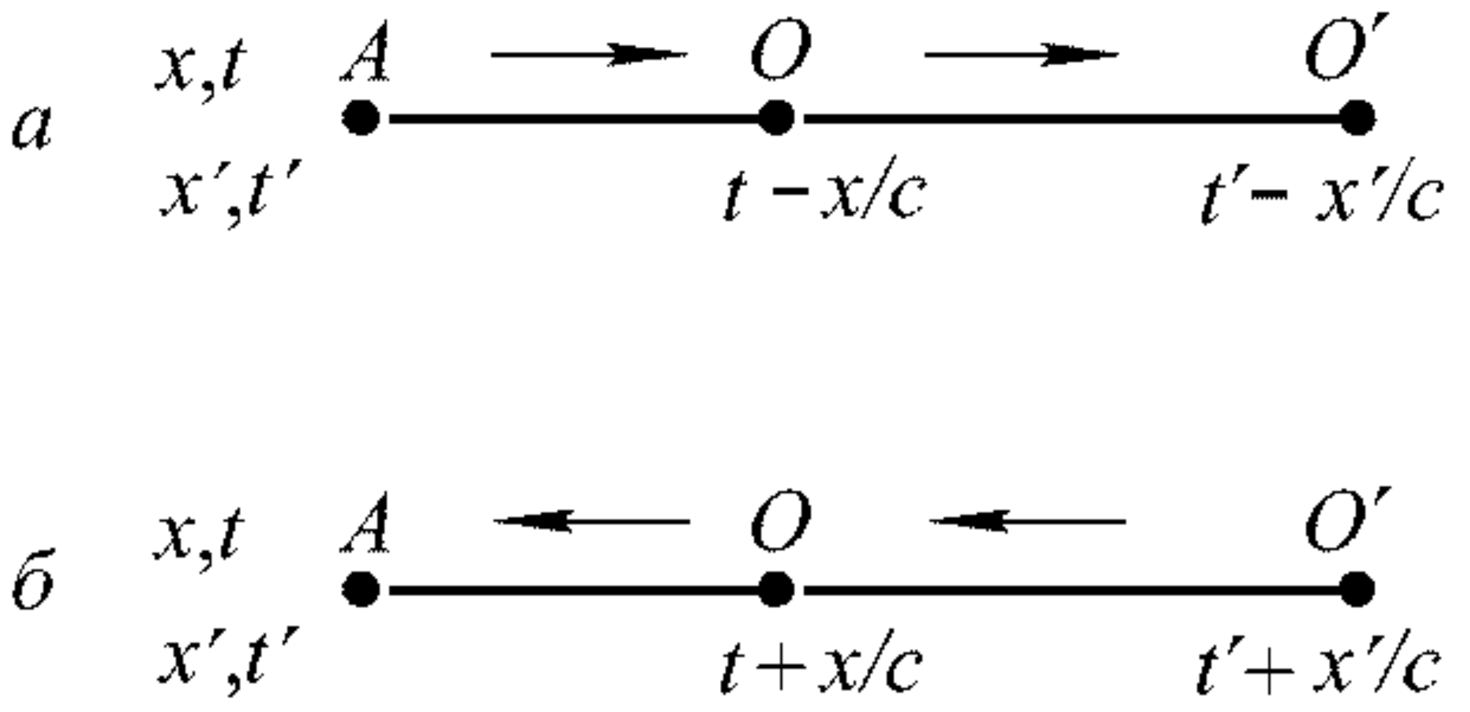


Рис. 330

Очень полезно рассмотреть все случаи взаимного расположения точек A, O, O' и убедиться, что во всех случаях справедливы соотношения (105.10).

Из соотношений (105.10) находим

$$\begin{aligned} x' &= \frac{1}{2}\left(k + \frac{1}{k}\right)x - \frac{1}{2}\left(k - \frac{1}{k}\right)ct, \\ t' &= \frac{1}{2}\left(k + \frac{1}{k}\right)t - \frac{1}{2}\left(k - \frac{1}{k}\right)\frac{x}{c}, \end{aligned} \tag{105.11}$$

или на основании (105.8)

$$x' = \frac{x - Vt}{\sqrt{1-\beta^2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - Vx/c^2}{\sqrt{1-\beta^2}}, \tag{105.12}$$

где введено обозначение

$$\beta = V/c. \tag{105.13}$$

Мы добавили формулы $y' = y$, $z' = z$, которые показывают, что поперечные координаты события y и z не преобразуются. Для доказательства последнего утверждения изготовим в системах отсчета S и S' одним и тем же способом два одинаковых твердых стержня, каждый из которых неподвижен в своей системе отсчета. Установим их своими концами на оси X параллельно осям Y и Y' (рис. 331). Длины стержней y и y', измеренные соответственно в системах S и S', конечно, будут одинаковы. Но это еще не означает, что справедливо второе уравнение (105.12). Надо еще показать, что y и y' можно рассматривать как координаты *одного и того же события*. Для доказательства этого к свободным концам стержней прикрепим маленькие шарики A и A'. Пусть сначала щарик A' расположен левее шарика A. Мы утверждаем, что при движении шарик A' обязательно столкнется с шариком A. Действительно, если бы движущийся шарик прошел выше или ниже неподвижного, то системы отсчета S и S' не были бы эквивалентны. Но если шарики столкнутся, то y и y' становятся координатами одного и того же события — столкновения шариков, и второе уравнение (105.12) может считаться доказанным. Так же доказывается и третье уравнение (105.12).

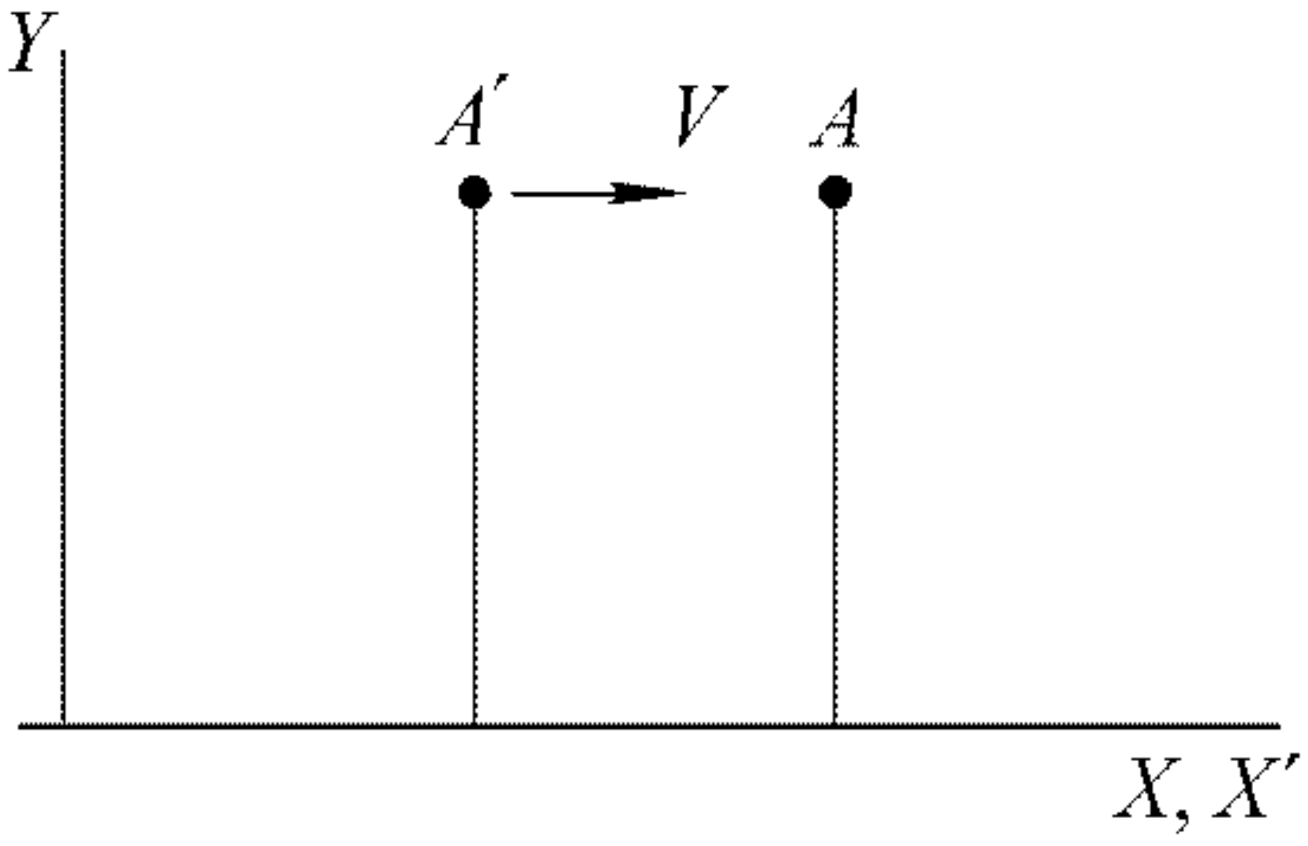


Рис. 331

Формулы (105.12) и решают задачу о преобразовании координат и времени при переходе от одной системы отсчета к другой. Они называются *преобразованием Лорентца* (этот термин был введен Пуанкаре). Лорентц получил их в 1904 г. К тем же формулам несколько раньше (в 1900 г.) пришел Лармор. И Лармор, и Лорентц, однако, принципиально стояли на точке зрения неподвижного эфира. У них истинным было только время t в системе отсчета, в которой эфир покоится. Величина же t' лишь формально играла роль времени — это была математическая переменная, вводимая таким образом, чтобы соблюдалась инвариантность уравнений электродинамики при переходе от переменных x, y, z, t к переменным x', y', z', t'. Настоящий вывод формул преобразования Лорентца и установление их истинного смысла дал Эйнштейн в 1905 г. В его теории все инерциальные системы отсчета совершенно эквивалентны, а t' является таким же «истинным временем», как и t. Это проявляется, в частности, если уравнения (105.12) разрешить относительно x, y, z, t. Таким путем получатся формулы «обратного преобразования»

$$x = \frac{x' + Vt}{\sqrt{1-\beta^2}}, \quad y = y', \quad z = z', \quad t = \frac{t' + Vx'/c^2}{\sqrt{1-\beta^2}}, \tag{105.14}$$

имеющие тот же вид, что и формулы «прямого преобразования» (105.12). Как и следовало ожидать, они получаются из формул (105.12) простой заменой V на $-V$.

Формулы (105.12) и (105.14) при $\beta > 1$ дали бы мнимые значения для координат и времени. Поэтому нет смысла говорить о движении одной системы отсчета со скоростью V, превышающей скорость света c. Отсюда следует, что скорость любого тела не может превышать c, так как с каждым телом можно связать систему отсчета.

При медленных движениях, когда $(V/c)^2 \ll 1$ и $Vv/c^2 \ll 1$ (v — скорость движения тела), преобразование Лорентца, как и следовало ожидать, в пределе переходит в преобразование Галилея.

4. В дорелятивистской физике пространство и время считались *независимыми* друг от друга. Расстояния между двумя пространственными точками (точнее, расстояния между двумя материальными точками в один и тот же момент времени), а также промежутки времени между двумя событиями считались *одинаковыми во всех системах отсчета*. Иными словами, обе эта величины считались *инвариантными* при переходе от одной системы отсчета к другой. В теории относительности такая инвариантность была утрачена. Вместо двух инвариантов — пространственного и временного — в ней сохранился только *один, пространственно-временной, инвариант*. Его легко найти, перемножая почленно уравнения (105.11). Это дает

$$c^2t^2 - x^2 = c^2t'^2 - x'^2 = \mathrm{Inv}\ . \tag{105.15}$$

Для удобства запишем этот инвариант через новые временны́е переменные

$$\tau = ct, \quad \tau' = ct', \tag{105.16}$$

имеющие размерность длины. Введение таких переменных означает, что промежутки времени теперь измеряются теми же единицами, что и пространственные расстояния: за единицу времени принимается время, в течение которого свет проходит единицу расстояния. В новых переменных

$$\tau^2 - x^2 = \tau'^2 - x'^2 = \mathrm{Inv}\ . \tag{105.17}$$

Учтем теперь, что при выбранной нами ориентации координатных осей $y' = y$, $z' = z$. Поэтому

$$\tau^2 - (x^2 + y^2 + z^2) = \tau'^2 - (x'^2 + y'^2 + z'^2) = \mathrm{Inv}\,. \qquad (105.18)$$

В этом виде пространственно-временной инвариант (105.18) уже не зависит от ориентации систем отсчета S и S' относительно друг друга, а также от направления скорости V. Однако в формуле (105.18) предполагается, что одно из событий фиксировано. Таким событием является совмещение начал координат O и O'. Чтобы освободиться от этого ограничения, запишем (105.18) в виде

$$\Delta\tau^2 - (\Delta x^2 + \Delta y^2 + \Delta z^2) = \Delta\tau'^2 - (\Delta x'^2 + \Delta y'^2 + \Delta z'^2) = \mathrm{Inv}\,, \qquad (105.19)$$

где $\Delta\tau$, Δx, ..., $\Delta\tau'$, $\Delta x'$, ... означают разности между временами и пространственными координатами двух событий 1 и 2 в системах отсчета S и S':

$$\Delta\tau = \tau_2 - \tau_1, \quad \Delta x = x_2 - x_1, \quad \Delta y = y_2 - y_1, \quad \Delta z = z_2 - z_1,$$
$$\Delta\tau' = \tau_2' - \tau_1', \quad \Delta x' = x_2' - x_1', \quad \Delta y' = y_2' - y_1', \quad \Delta z' = z_2' - z_1'.$$

Квадратный корень из инварианта (105.19) называется *интервалом между рассматриваемыми событиями* и в дальнейшем обозначается через s_{12} или Δs. Очевидно, квадрат интервала между событиями 1 и 2 можно представить в виде

$$s_{12}^2 = (\tau_2 - \tau_1)^2 - l_{12}^2 = (\tau_2' - \tau_1')^2 - l_{12}'^2, \qquad (105.20)$$

где l_{12} и l_{12}' — расстояния между точками, в которых произошли события, в системах S и S' соответственно.

Можно также в формулы преобразования Лорентца ввести параметры τ и τ' вместо t и t'. Ограничиваясь частным случаем, представленным на рис. 328, перепишем формулы (105.12) в виде

$$x' = \frac{x - \beta\tau}{\sqrt{1-\beta^2}}, \qquad \tau' = \frac{\tau - \beta x}{\sqrt{1-\beta^2}}. \qquad (105.21)$$

Величина β имеет смысл скорости движущейся системы координат, если за единицу принять скорость света.

Минковский (1864–1909) для описания пространственно-временных событий ввел геометрическую терминологию. Совокупность значений τ, x, y, z, характеризующую время и место события, он назвал *мировой точкой*. Многообразие мировых точек есть четырехмерное пространство, называемое *миром* или *пространством Минковского*. Линия в пространстве Минковского называется *мировой линией*. Интервал между двумя событиями принимается за *инвариантное расстояние* между соответствующими мировыми точками. На основе таких представлений было создано тензорное исчисление в пространстве Минковского, аналогичное тензорному исчислению обычной эвклидовой геометрии. Оно является адекватным математическим аппаратом специальной теории относительности.

5. Интервалы между событиями можно разделить на *вещественные* и *чисто мнимые*. Ввиду инвариантности интервала, это деление *не зависит от выбора системы отсчета.*

Поставим вопрос, можно ли выбрать такую систему отсчета, в которой события 1 и 2 были бы *одноместны*, т.е. происходили бы в одной и той же точке пространства? Если S' — такая система, то в ней $l'_{12} = 0$, и на основании (105.20) квадрат интервала может быть представлен в виде $s^2_{12} = (\tau'_2 - \tau'_1)^2$. Отсюда видно, что $s^2_{12} > 0$, т.е. необходимо, чтобы интервал s_{12} был вещественным. Для доказательства достаточности этого условия можно без нарушения общности ограничиться частным преобразованием Лорентца (105.21). Чтобы рассматриваемые события в системе S' пространственно совпадали, достаточно, чтобы выполнялось условие $\Delta x' = 0$, т.е. $\Delta x = \beta\,\Delta\tau$. Отсюда видно, что система S' должна двигаться со скоростью $\beta = \Delta x/\Delta\tau$. Но для вещественных интервалов $|\Delta x| < \Delta\tau$, так что $|\beta| < 1$. Значит, система S' должна двигаться со скоростью, меньшей скорости света, а потому ее можно реализовать. Промежуток времени между одноместными событиями в системе отсчета S' будет равен $\Delta\tau' = |s_{12}|$, или в обычных единицах $\Delta t' = |s_{12}|/c$. Вещественные интервалы называются *времениподобными.*

Поставим теперь вопрос о существовании системы отсчета, в которой события 1 и 2 были бы *одновременны*. Если S' — такая система, то $\tau'_2 - \tau'_1 = 0$ и, следовательно, на основании (105.20) должно быть $s^2_{12} = -l'^2_{12}$. Значит, необходимо, чтобы интервал s_{12} был чисто мнимым. Достаточность этого условия доказывается совершенно так же, как в предыдущем случае. Расстояние между точками, в которых произошли одновременные события 1 и 2, в системе S' равно $l'_{12} = |s_{12}|$. Чисто мнимые интервалы называются *пространственноподобными.*

Рассмотрим, наконец, особый случай, когда интервал между событиями равен нулю. В этом случае, чтобы сделать события одноместными, надо перейти к системе отсчета, движущейся со скоростью света. То же требуется, чтобы события стали одновременными. И то и другое невозможно. Нулевые интервалы называются *световыми.* Такими интервалами связаны отправление светового сигнала из некоторой точки и приход его в другую точку пространства.

6. В случае частного преобразования Лорентца (105.21), соответствующего рис. 328, все изложенное можно наглядно интерпретировать графически. Это возможно потому, что достаточно ограничиться рассмотрением преобразования только одной пространственной координаты x и времени τ, т.е. рассуждать так, как если бы пространство Минковского было *двумерной плоскостью* (x, τ). Произвольное событие (мировую точку) O в этой плоскости примем за начало прямоугольной системы координат (рис. 332). Проведем через O взаимно перпендикулярные оси, одну из которых примем за пространственную ось x, а другую — за временную ось τ в системе отсчета S. Штриховые прямые $\tau = x$ и $\tau = -x$ будут мировыми линиями световых сигналов, распространяющихся соответственно в положительном и отрицательном направлениях оси X. Пусть A — произвольное событие. Соединим точки A и O прямой мировой линией $\tau = x \operatorname{tg} \alpha$, описывающей равномерное движение какого-то тела вдоль оси X со скоростью β. Если $\beta < 1$, то с этим телом можно связать систему отсчета S', а прямую OA принять за пространственную ось X'. Ось времени τ' найдется из условия, что на ней $x' = 0$. Как видно из (105.21), это будет мировая линия $x = \beta\tau$, т.е. прямая OB, наклоненная под тем же углом $\alpha = \operatorname{arctg} \beta$, но уже к временной оси τ. В новой системе отсчета события O и A будут одновременными, но не одноместными. События же O и B одноместны, но не одновременны.

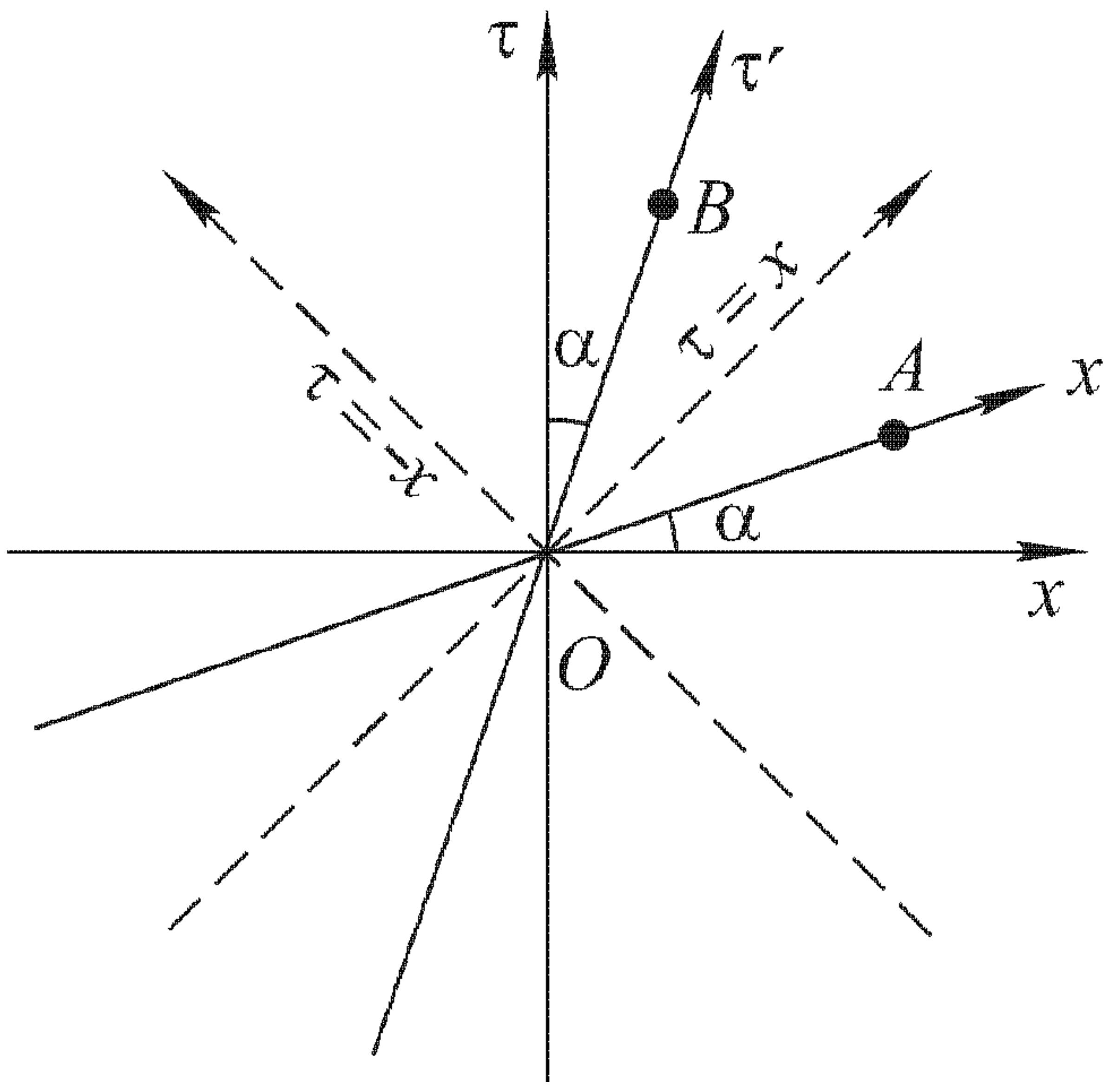


Рис. 332

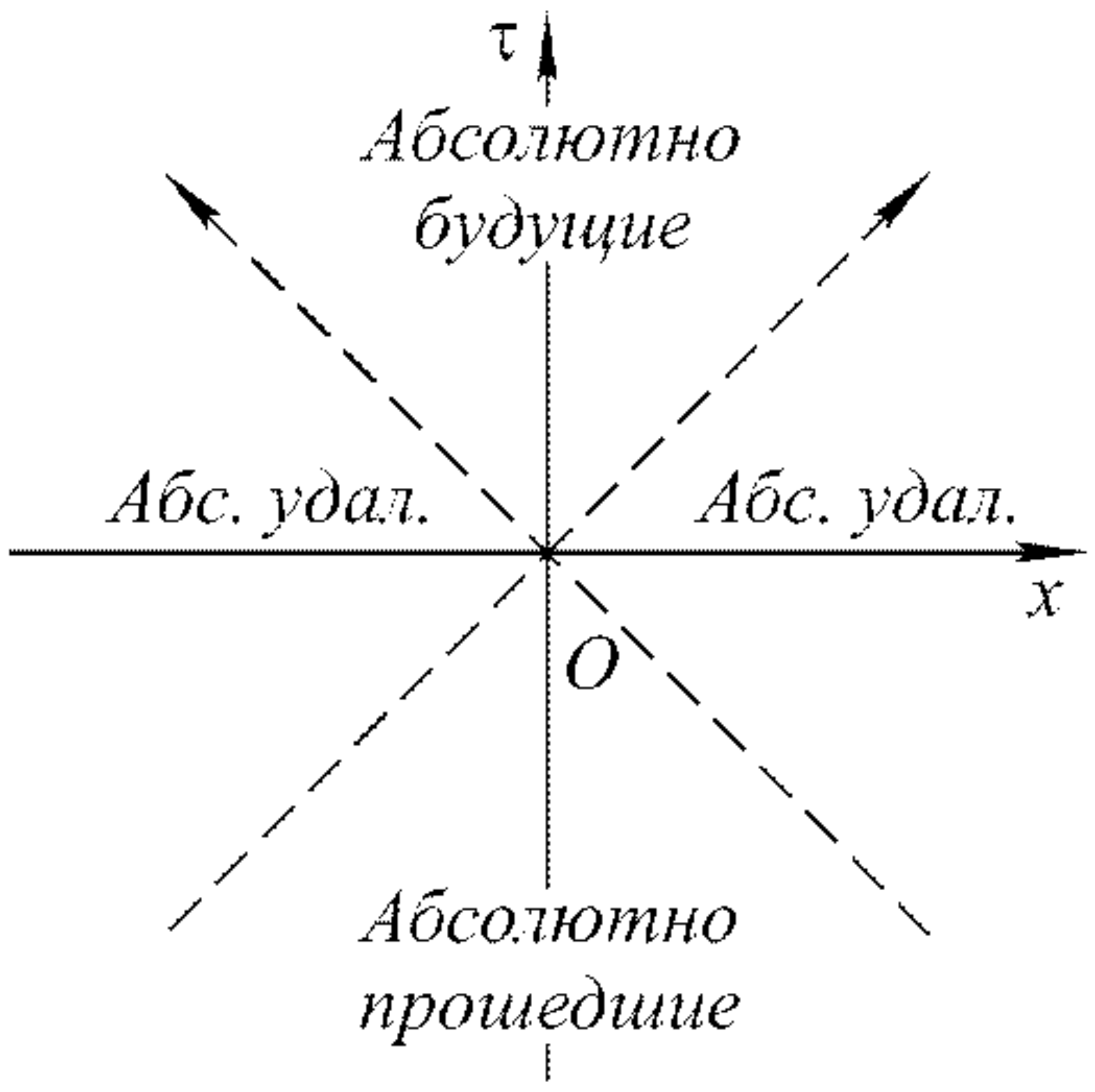


Рис. 333

Вообще, все события по отношению к событию O можно разделить на *абсолютно удаленные* и *абсолютно неодновременные* (т.е. *абсолютно прошедшие* и *абсолютно будущие*), как это показано на рис. 333. Для первых интервал между событиями чисто мнимый, для вторых — вещественный. Границей между такими событиями служат штриховые мировые линии световых сигналов, распространяющихся в положительном и отрицательном направлениях пространственной оси. В четырехмерном пространстве Минковского такой границей будет трехмерное многообразие, а именно конус $\tau^2 - x^2 - y^2 - z^2 = 0$, осью которого является ось времени τ. Он называется *световым конусом.*

Причинно связанными могут быть только события, интервал между которыми времениподобный. Например, событие O (см. рис. 332) могло бы быть причиной события B, так как в любой системе отсчета событие B наступает позже события O. Но события не могут быть причинно связанными, когда интервал между ними пространственноподобный. Таковы, например, события O и A (рис. 332). В неподвижной системе отсчета S событие A происходит позже события O. В штрихованной системе (x', τ') оба эти события одновременны. Если же взять систему отсчета, движущуюся быстрее системы (x', τ'), но все еще медленнее света, то ее пространственная ось будет на рис. 332 наклонена круче оси X'. В такой системе отсчета событие A произойдет раньше события O. Таким образом, нельзя удовлетворить требованию, чтобы в любой системе отсчета «причина» предшествовала «следствию». Это и доказывает наше утверждение.

Прямая OA с уравнением $\tau = \beta x$ есть мировая линия некоторого движения, происходящего со скоростью $x/\tau = 1/\beta$, т.е. со *сверхсветовой скоростью.* Существование сверхсветовых скоростей не противоречит теории относительности. Последняя допускает любые скорости. Однако в случае распространения состояний со сверхсветовыми скоростями интервал между любыми двумя состояниями будет пространственноподобным, а потому каждое из этих состояний не может быть причиной другого. Такие процессы не могут служить «сигналами» для передачи информации. *Все тела и сигналы, передающие воздействие, не могут распространяться со скоростью, превышающей скорость света в вакууме. Скорость света в вакууме есть максимально возможная скорость распространения взаимодействий* [1].

§ 106. Лорентцово сокращение длины и замедление времени

1. Если твердый стержень покоится в какой-то системе отсчета, то его длина l_0 определяется сравнением с масштабным стержнем, покоящимся в той же системе отсчета. Величину l_0 можно назвать *собственной длиной стержня*, поскольку она не зависит от выбора системы отсчета, в которой покоится стержень. Но если стержень движется, то необходимо условиться, что понимать под его длиной в покоящейся системе отсчета. Во избежание недоразумений специально подчеркнем, что все измерения расстояний и промежутков времени во всякой системе отсчета должны производиться с помощью масштабных стержней и часов, *неподвижных* в этой системе.

---

[1] За последние примерно 10 лет обсуждается вопрос о существовании гипотетических сверхсветовых частиц — *тахионов.* Ясно, что если бы такие частицы существовали, то они не могли бы служить «сигналами» для передачи взаимодействий. Подробнее см. *Угаров В.А.* Специальная теория относительности.—М.: Наука, 1977.

Длиной l движущегося стержня в покоящейся системе отсчета называется расстояние между двумя точками в этой системе, мимо которых концы стержня проходят одновременно. Для нахождения связи между l и l_0 воспользуемся частной формой преобразования Лорентца (105.12). Пусть стержень покоится в системе S' и лежит на оси X'. Тогда разность координат его концов $\Delta x'$ в системе S' и будет длиной l_0 покоящегося стержня. Разность же координат тех же концов Δx в системе S, взятая в один и тот же момент t, будет длиной l движущегося стержня. Но из первой формулы (105.12) при $t = \mathrm{const}$ следует $\Delta x' = \Delta x/\sqrt{1-\beta^2}$, а потому

$$l = l_0\sqrt{1-\beta^2}. \qquad (106.1)$$

Таким образом, *длина движущегося стержня короче, чем покоящегося.* Это явление называется *лорентцовым сокращением длины.* Первоначально оно было введено независимо друг от друга Фицжеральдом (1851–1901) и Лорентцом. Но у них это была гипотеза *ad hoc*, специально придуманная для объяснения отрицательного результата опыта Майкельсона, хотя Лорентц и пытался обосновать ее с точки зрения электронной теории. В теории относительности лорентцово сокращение получается, а его истинный физический смысл устанавливается без каких бы то ни было добавочных предположений. Это, конечно, не исключает возможности атомистического объяснения лорентцова сокращения, а также явления замедления хода движущихся часов, о котором говорится дальше. Но для этого надо располагать не только уравнениями электродинамики, но и пока еще не известными законами, определяющими строение вещества. Когда такие законы будут установлены, можно на их основе рассмотреть и вопрос об изменении длин и времен. Однако если теория относительности верна, то результат такого рассмотрения заранее известен. Правильные законы природы должны быть *инвариантны относительно преобразования Лорентца*, а потому в вопросе о сокращении длин и замедлении времени они не могут привести к иным результатам, чем теория относительности.

2. Рассмотрим теперь какие-либо два события, интервал между которыми времениподобный. Промежутки времени Δt и $\Delta t'$ между этими событиями, измеренные в «неподвижной» S и «движущейся» S' системах отсчета, вообще говоря, будут разными. Конечно, обе системы S и S' совершенно равноправны, так что между Δt и $\Delta t'$ может существовать любое соотношение $\Delta t \lesseqgtr \Delta t'$. Допустим теперь, что в качестве движущейся системы S' взята такая, в которой оба события происходят *в одном и том же месте пространства.* (В случае времениподобных интервалов такая система существует, см. п. 5 предыдущего параграфа.) Тогда система S' становится выделенной среди всех прочих инерциальных систем отсчета. Время, измеренное в такой системе, будем обозначать через t_0, а в «неподвижной» системе S — по-прежнему через t. Для нахождения соотношения между Δt и Δt_0 будем предполагать, что в (105.14) координата x' постоянна, т.е. одна и та же для обоих событий. Тогда из последней формулы (105.14) получим $\Delta t = \Delta t'/\sqrt{1-\beta^2}$. Но, согласно нашему определению, $\Delta t'$ и есть Δt_0. Следовательно,

$$\Delta t_0 = \Delta t\sqrt{1-\beta^2}. \tag{106.2}$$

Из (106.2) видно, что $\Delta t_0 < \Delta t$, т.е. *промежуток времени между двумя событиями минимален в той системе отсчета, в которой эти события одноместны.* Это явление называется *замедлением хода движущихся часов.* Причина такого названия заключается в следующем. Допустим, что взяты какие-то одни часы, помещенные в одну из точек системы S'. Такие часы называются «движущимися», поскольку они движутся вместе с системой S'. Все же часы, покоящиеся в системе S, называются «неподвижными». Ясно, что в системе S' любые два события, происходящие в месте нахождения «движущихся» часов, будут одноместны, так что эти часы измерят промежуток времени Δt_0. Здесь моменты наступления обоих событий отмечаются по *одним и тем же* («движущимся») часам, так что никакой синхронизации не требуется. Напротив, в «неподвижной» системе S моменты наступления событий отмечаются по часам, находящимся в *различных местах пространства.* Для измерения Δt необходимо иметь *двое часов*, синхронизованных между собой по правилу Эйнштейна. С этим и связана *асимметрия* при измерении времени по «неподвижным» и «движущимся» часам, проявляющаяся в неравенстве $\Delta t_0 < \Delta t$.

3. Допустим теперь, что какая-либо частица движется относительно «неподвижной» системы отсчета S по криволинейной траектории с переменной по величине скоростью v. В специальной теории относительности допускаются только такие пространственно-временные системы отсчета, которые движутся относительно S равномерно и прямолинейно. Возьмем бесконечное множество таких систем, движущихся со всевозможными скоростями и во всевозможных направлениях. Система отсчета, относительно которой мгновенная скорость частицы равна нулю, и связанные с ней часы называются *сопутствующими.* При движении частица непрерывно переходит из одной сопутствующей системы отсчета в другую. Разобьем траекторию частицы в системе S на бесконечно короткие отрезки. Пусть dt — время, затрачиваемое в системе S на прохождение одного из таких отрезков. Согласно (106.2), по сопутствующим часам на то же движение потребуется время $dt_0 = dt\sqrt{1-v^2/c^2}$. Конечный промежуток времени, измеренный по неподвнжным часам, представится интегралом

$$t_2 - t_1 = \int \frac{dt_0}{\sqrt{1-v^2/c^2}}, \tag{106.3}$$

взятым по всей траектории движения.

Если часы движутся вместе с частицей, то возникает вопрос, как связана длительность, измеренная по таким часам, с длительностью, измеренной по сопутствующим часам? На этот вопрос нельзя ответить *без рассмотрения конкретного устройства часов*, так как при ускоренном движении появляются силы инерции, влияющие на ход часов. Например, в случае движения пружинных часов силы инерции могут деформировать балансир или другие части их. Деформации могут быть настолько большими, что часы остановятся или сломаются. Маятниковые часы вообще станут непригодными для измерения времени в отсутствие поля тяготения. Сердце живого организма (а оно в принципе может рассматриваться тоже как «часы») перестанет биться, а сам организм погибнет, если ему сообщить достаточно большое ускорение, и т.д. Однако в принципе допустимы и такие часы, на ход которых сила тяжести и силы инерции оказывают ничтожное влияние, в пределе совсем исчезающее. Такие часы условимся называть *идеальными.* Измеряемая ими длительность совпадает с длительностью, измеренной по сопутствующим часам. Этот признак в принципе позволяет экспериментально проверить, являются ли те или иные конкретные часы идеальными или не являются. Ясно, что в формуле (106.3) под t_0 можно понимать время, отсчитываемое *по идеальным часам, движущимся вместе с частицей.* Это время называется *собственным временем*, поскольку оно инвариантно, т.е. не зависит от движения частицы. Из (106.3) видно, что длительности по «неподвижным» и «движущимся» идеальным часам связаны соотношением

$$t_2 - t_1 > t_{02} - t_{01}. \tag{106.4}$$

В частности, если часы вернутся в исходное положение (относительно системы S), то они покажут время, меньшее времени по «неподвижным» часам.

4. Наилучшим приближением к идеальным часам являются *атомные* и в особенности *ядерные часы*. Силы инерции по своему действию эквивалентны гравитационным силам (см. т. I, § 71). В обычных условиях как те, так и другие силы пренебрежимо малы по сравнению с электрическими и еще более мощными ядерными силами, определяющими процессы в электронных оболочках и ядрах атомов. В таких процессах гравитационные силы и силы инерции практически не играют никакой роли.

Рассмотрим, например, атом цезия-133 в *цезиевых эталонных часах*, с помощью которых устанавливается эталон времени — *секунда*. По принятому соглашению электромагнитное излучение такого атома в отсутствие внешних полей совершает $\nu = 9\,192\,631\,770 \approx 10^{10}$ колебаний в секунду. Радиус атома $\sim 10^{-8}$ см. Если атом уподобить гармоническому осциллятору, то при колебаниях будут развиваться громадные ускорения $\sim (2\pi\nu)^2 r \sim 10^{13}$ см/с^2. Если часам сообщить ускорение ~ 1 см/с^2, то для этого потребуются силы в 10^{13} раз слабее. Ясно, что никакого существенного действия на процессы внутри атома они оказать не могут. Только при измерениях времени с относительной точностью $\sim 10^{-12}$–10^{-13} такие ускорения и соответствующие им гравитационные поля, возможно, могут сказаться на ходе часов и результатах измерений. Тогда их надо учитывать. Из сказанного ясно, что из всех часов атомные и ядерные часы являются, пожалуй, единственными, которые пригодны для обнаружения и исследования релятивистского замедления времени, по крайней мере при скоростях, малых по сравнению со скоростью света.

Явление радиоактивного распада атомов или других нестабильных частиц может выполнять роль идеальных часов. Радиоактивный распад подчиняется закону

$$n = n_0 e^{-t/\tau}, \tag{106.5}$$

где n_0 — начальное число частиц, а n — число их через время t. Постоянная τ называется *временем жизни* рассматриваемой нестабильной частицы. О времени t можно судить по отношению n/n_0. Релятивистское замедление времени было подтверждено в явлении распада *мюонов* (μ-мезонов). Так называются нестабильные заряженные частицы, масса которых в 207 раз превышает массу электрона. Заряд мюона равен заряду электрона, но может быть и положительным, и отрицательным. Мюоны образуются в космических лучах в верхних слоях атмосферы (на высотах порядка 10 км). Сравнение интенсивностей потока мюонов в космических лучах на горе и у ее основания показало, что среднее время жизни мюона в лабораторной системе отсчета $\tau \approx 10^{-5}$ с. С другой стороны, космические мюоны можно было замедлить в свинцовом блоке и с помощью специального устройства отфильтровать медленные мюоны. Измерения показали, что время жизни медленного (покоящегося) мюона $\tau_0 \approx \approx 2{,}20 \cdot 10^{-6}$ с. Если бы не было релятивистского замедления времени, то поток космических мюонов, даже если бы они двигались со скоростью света, уменьшался бы в e раз при прохождении расстояния ≈ 600 м. На расстоянии уже 5 км их интенсивность уменьшилась бы в $e^{50/6} \approx 4000$ раз, т.е. мюоны вообще не могли бы достигать поверхности земли. В действительности их интенсивность при прохождении такого расстояния уменьшается примерно в $e^{5/3} \approx 5$ раз. Учет релятивистского замедления времени устраняет это противоречие. Действительно, время жизни мюона в лабораторной системе отсчета τ связано с собственным временем жизни соотношением $\tau = \tau_0/\sqrt{1-\beta^2}$. Измерение средней кинетической энергии космических мюонов показало, что она $\approx 10^9$ эВ. По этой энергии нетрудно рассчитать $\sqrt{1-\beta^2} \approx 0{,}1$. Поэтому следует ожидать, что $\tau \approx 2{,}2 \cdot 10^{-6}/0{,}1 \approx \approx 2 \cdot 10^{-5}$ с. Это по порядку величины согласуется со значением, полученным на опыте.

5. С построением мощных ускорителей заряженных частиц опыты подобного рода производились в более определенных и лучше контролируемых условиях. Наиболее подходящими частицами для таких опытов оказались заряженные *пионы* (иначе называемые π-мезонами). Их масса в 273 раза больше массы электрона, а заряд равен заряду электрона. Пионы во множестве образуются при взаимодействии протонов высоких энергий с веществом. Среднее время жизни пиона в системе отсчета, где он покоится, $\tau_0 = 2{,}60 \cdot 10^{-8}$ с. На циклотроне Колумбийского университета был получен пучок пионов со скоростью $v = 0{,}75\,c$ ($\beta = 0{,}75$). Если бы не было релятивистского замедления времени, то за время τ_0 пучок проходил бы расстояние $0{,}75 \cdot 3 \cdot 10^{10} \times \times 2{,}60 \cdot 10^{-8} = 5{,}85$ м. На самом деле, как показали измерения, расстояние, на котором интенсивность пучка уменьшается в e раз, равно $8{,}5 \pm 0{,}6$ м, т.е. в лабораторной системе время жизни пиона τ в $1{,}45 \pm 0{,}11$ раза больше собственного времени жизни τ_0. Но это хорошо согласуется с формулой $\tau/\tau_0 = 1/\sqrt{1-\beta^2}$, которая дает $\tau/\tau_0 = 1{,}51$.

Результат опыта можно интерпретировать и как проявление *релятивистского сокращения длины*. Действительно, в системе отсчета, где пион покоится, его время жизни равно τ_0. Лаборатория движется относительно пиона со скоростью v. За время

τ_0 она проходит в системе пиона расстояние $l = v\tau_0$. Если то же расстояние измерять масштабным стержнем, который покоится в лаборатории, то оно окажется равным $l_0 = l/\sqrt{1-\beta^2} = 8{,}8$ м, что согласуется с опытом.

6. Хафель и Китинг поставили опыт для обнаружения релятивистского замедления хода часов уже в *макроскопических условиях*. В этом опыте четыре экземпляра цезиевых часов в октябре 1971 г. были помещены на реактивных самолетах, облетевших вокруг земного шара в восточном и западном направлениях. Временные интервалы $t_{\text{вост}}$ и $t_{\text{зап}}$, измеренные по часам, двигавшимся соответственно на восток и на запад, сравнивались с интервалами $t_{\text{неп}}$, измеренными эталонными неподвижными часами, находившимися в Морской обсерватории в Вашингтоне. После усреднения по четырем движущимся часам получились такие результаты:

$$t_{\text{вост}} - t_{\text{неп}} = (-59 \pm 10)\cdot 10^{-9}\ \text{с}, \qquad t_{\text{зап}} - t_{\text{неп}} = (273 \pm 7)\cdot 10^{-9}\ \text{с}.$$

Посмотрим, что следует ожидать согласно теории относительности. Ускорение центра Земли, вызванное гравитационным притяжением Солнца, составляет примерно 0,18 ускорения, которое получают точки земного экватора из-за осевого вращения Земли. Хотя это ускорение и немало, его можно не принимать во внимание, ввиду *принципа эквивалентности сил тяготения и сил инерции* (см. т. I, § 71). Если не учитывать неоднородность гравитационного поля Солнца, то это поле будет полностью компенсировано центробежной силой инерции, обусловленной вращением центра Земли вокруг Солнца. Таким образом, можно считать, что относительно инерциальной системы отсчета центр Земли движется прямолинейно и равномерно. Поэтому невращающаяся система отсчета с началом координат в центре Земли практически будет также инерциальной системой отсчета. Ее мы и используем в последующих рассуждениях. Пусть самолеты облетают земной шар по параллели со скоростью v относительно Земли на постоянной высоте h. Их скорости относительно инерциальной системы отсчета будут $V_{\text{вост}} = V_0 + v$, $V_{\text{зап}} = V_0 - v$, где V_0 — скорость наземной лаборатории в той же системе. Сначала будем считать, что $h = 0$. Тогда, если t_0 — собственное время, то

$$t_{\text{неп}} = \int dt_0 \Big/ \sqrt{1 - V_0^2/c^2} \approx t_0 - \frac{V_0^2}{2c^2}\, t_0.$$

Аналогично,

$$t_{\text{вост}} \approx t_0 - \frac{(V_0+v)^2}{2c^2}\, t_0, \qquad t_{\text{зап}} \approx t_0 - \frac{(V_0-v)^2}{2c^2}\, t_0.$$

Вычтем отсюда предыдущие выражения и учтем, что в окончательных формулах в пределах точности расчета время t_0 можно заменить на $t_{\text{неп}}$. Тогда получим

$$t_{\text{вост}} - t_{\text{неп}} = -\frac{2V_0 v + v^2}{2c^2}\, t_{\text{неп}}, \quad t_{\text{зап}} - t_{\text{неп}} = \frac{2V_0 v - v^2}{2c^2}\, t_{\text{неп}}. \tag{106.6}$$

Естественно, под $t_{\text{неп}}$ надо понимать время нахождения самолетов в воздухе *без учета времени остановок*, так как время остановок не влияет на величину разностей $t_{\text{вост}} - t_{\text{неп}}$ и $t_{\text{зап}} - t_{\text{неп}}$.

Допустим, что самолет облетает земной шар в течение суток. Тогда для широты Вашингтона потребуется скорость $v = V_0 \approx 1000$ км/ч ≈ 300 м/с. По формулам (106.6) найдем $t_{\text{вост}} - t_{\text{неп}} \approx -130 \cdot 10^{-9}$ с, $t_{\text{зап}} - t_{\text{неп}} \approx +43 \cdot 10^{-9}$ с.

В приведенном вычислении не учтено влияние *гравитационного потенциала Земли* на течение времени (см. § 109, а также т. I, § 72). С учетом этого обстоятельства оба выражения (106.6) надо увеличить на

$$\Delta t_{\text{грав}} = \frac{gh}{c^2}\, t_{\text{неп}}, \tag{106.7}$$

где h — высота полета, а g — ускорение свободного падения на поверхности Земли. Если положить $h = 10$ км, то в приведенном выше примере получится $\Delta t_{\text{грав}} = 94 \cdot 10^{-9}$ с, т.е. «гравитационный эффект» того же порядка, что и учтенный выше «кинематический эффект». Хафель и Китинг провели вычисление с учетом этой поправки, выполнив интегрирование по фактически совершенным маршрутам самолетов. Они получили теоретические значения:

$$t_{\text{вост}} - t_{\text{неп}} = (-40 \pm 23) \cdot 10^{-9} \text{ с}, \quad t_{\text{зап}} - t_{\text{неп}} = (275 \pm 21) \cdot 10^{-9} \text{ с},$$

удовлетворительно согласующиеся с полученными на опыте. Заметим, что из разности $(t_{\text{зап}} - t_{\text{вост}})$ поправка $\Delta t_{\text{грав}}$ исключается. Эта разность обусловлена только «кинематическим эффектом» замедления времени. Совпадение теоретических результатов с измеренными на опыте рассматривается как доказательство не вызывающего сомнения релятивистского замедления хода движущихся часов.

7. Остановимся в заключение на так называемом *парадоксе близнецов*. Из двух братьев-близнецов A остается на Земле, а B отправляется в кругосветное путешествие на межзвездном корабле, двигаясь со скоростью, близкой к скорости света. Через 5 лет по своим часам брат B возвращается обратно и находит брата A глубоким стариком. Оказалось, что за время путешествия по часам на Земле прошло 50 лет. Таким образом, открывается возможность за время человеческой жизни совершать не только путешествия к далеким звездным мирам, но и *путешествия в будущее*. Если отвлечься от технической и практической стороны вопроса, то такие путешествия принципиально возможны.

В самом деле, биологические процессы не представляют собой какую-то обособленную группу явлений природы. Как и прочие явления природы, они подчиняются законам физики. Если на межзвездном корабле создать условия, близкие к условиям на Земле, то и жизненные процессы на нем будут протекать практически так же, как и на Земле. Биения сердца в человеческом организме выполняют роль часов. Если за время жизни сердце человека на Земле совершает $2 \cdot 10^9$ ударов, то столько же ударов оно совершит и на корабле. Но движущиеся часы идут медленнее неподвижных. Если за время путешествия сердце брата B совершит $1{,}5 \cdot 10^8$ ударов, то на Земле к моменту встречи сердце брата A успеет совершить ударов в 10 раз больше. Но это и есть «парадокс близнецов».

В 1974 г. парадокс близнецов был подтвержден экспериментально на ускорителе в ЦЕРНе (Европейский центр по ядерным исследованиям, Швейцария). Ускоренные мю-мезоны удерживались магнитным полем на круговой орбите радиуса 5 м в течение свыше 150 мкс. За это время они совершали более 10^5 оборотов. Энергия мезонов превышала энергию покоя примерно в 12 раз, так что $1/\sqrt{1-\beta^2} = 12$. Поэтому ожидаемое время жизни мезона в лабораторной системе должно составлять $2{,}2 \cdot 12 = 26{,}4$ мкс. Опыт дал для этого времени $26{,}37 \pm 0{,}05$ мкс.

§ 107. Эффект Доплера и аберрация света

1. Если в «неподвижной» системе S распространяется монохроматическая волна с определенной частотой ω и в определенном направлении, то в «движущейся» системе S' та же волна будет иметь другую частоту ω' и распространяться в другом направлении. Изменение частоты волны при переходе от одной системы отсчета к другой называется *эффектом Доплера* (1803–1853), а изменение направления — *аберрацией света*. При изложении теории этих явлений мы ограничимся случаем, когда световая волна распространяется *в вакууме*. Тогда в обеих системах отсчета пространство, в котором распространяется волна, будет *изотропным*. При наличии же среды изотропия сохранится только в той системе отсчета, в которой среда покоится. В системе отсчета, относительно которой среда движется, появляется анизотропия, вносящая осложнения в теорию эффекта Доплера и аберрации в материальных средах.

В релятивистской теории изменение частоты и направления распространения волны проще всего можно определить из условия *равенства фаз* одной и той же волны в обеих системах отсчета. Предполагая, что волна плоская, это условие можно записать в виде

$$\omega t - \mathbf{kr} = \omega' t' - \mathbf{k}'\mathbf{r}', \tag{107.1}$$

где $\mathbf{r}$, t и $\mathbf{r}'$, t' — координаты и время одного и того же события в рассматриваемых системах отсчета, причем начало отсчета времени выбрано так, что в момент совмещения начал координат обеих систем $t = t' = 0$.

Для доказательства сформулированного условия предположим, что в началах координат в момент $t = t' = 0$ на волне сделана метка, переносимая волной с фазовой скоростью, так что фаза колебания в месте нахождения метки в обеих системах отсчета будет $\varphi_1 = \varphi_1' = 0$. Вообразим теперь произвольно движущихся наблюдателей A и A', из которых A для определения своего положения в пространстве и времени пользуется координатами и часами системы S, а A' — координатами и часами системы S'. Можно, например, не нарушая общности, предположить, что наблюдатель A покоится в системе отсчета S, а наблюдатель A' — в системе отсчета S'. Однако это предположение в такой специальной форме не обязательно. (Обычно при рассмотрении эффекта Доплера систему S связывают с источником, а систему S' — с наблюдателем.) Оба наблюдателя начинают счет проходящих мимо них волн с того момента, когда мимо них проходит сделанная метка. Пусть оба наблюдателя встречаются друг с другом. Обозначим через $\mathbf{r}$, t координаты и время этого события в системе S, а через $\mathbf{r}'$, t' — в системе S'. Наблюдатель A найдет, что к моменту встречи фаза пришедшего к нему колебания будет $\varphi_2 = \omega t - \mathbf{kr}$, и сосчитает к этому моменту

$$N = \frac{\varphi_2 - \varphi_1}{2\pi} = \frac{\omega t - \mathbf{kr}}{2\pi}$$

волн. Наблюдатель A' найдет, что к моменту встречи мимо него прошло

$$N' = \frac{\omega' t' - \mathbf{k}'\mathbf{r}'}{2\pi}$$

волн. Но число прошедших волн не может зависеть от того, кто их считает, а потому

$$N = N', \qquad \text{или} \qquad \omega t - \mathbf{kr} = \omega' t' - \mathbf{k}'\mathbf{r}',$$

что и требовалось доказать.

Для нахождения формул, определяющих преобразование частоты и направления распространения волны в релятивистской теории, достаточно в формуле (107.1) выразить переменные $\mathbf{r}(x, y, z)$ и t через переменные $\mathbf{r}'(x', y', z')$ и t', воспользовавшись для этого формулами преобразования Лоренца (105.14). Сделав это и сравнив коэффициенты при одинаковых переменных, получим

$$\omega' = \frac{\omega - k_x V}{\sqrt{1-\beta^2}}, \tag{107.2}$$

$$k'_{x'} = \frac{k_x - \omega V/c^2}{\sqrt{1-\beta^2}}, \quad k'_{y'} = k_y, \quad k'_{z'} = k_z. \tag{107.3}$$

Аналогично получаются формулы обратного преобразования:

$$\omega = \frac{\omega' + k'_{x'} V}{\sqrt{1-\beta^2}}, \tag{107.4}$$

$$k_x = \frac{k'_{x'} + \omega V/c^2}{\sqrt{1-\beta^2}}, \quad k_y = k'_{y'}, \quad k_z = k'_{z'}. \tag{107.5}$$

Формулы (107.2) и (107.4) описывают эффект Доплера, а формулы (107.3) и (107.5) — аберрацию света.

2. Рассмотрим частный случай, когда волна распространяется вдоль оси X. В этом случае полагаем в формуле (107.2) $k_x = \omega/c$ и получаем

$$\omega' = \omega\sqrt{\frac{1-\beta}{1+\beta}}. \tag{107.6}$$

Изменение частоты в этом случае называется *продольным эффектом Доплера*. Понятно, что результат (107.6) можно получить и из общей формулы (107.4). Частота волны при наблюдении из системы S' уменьшается, когда эта система движется в направлении распространения волны, и увеличивается, когда она движется навстречу волне. При медленных движениях, когда можно пренебречь квадратами отношения $\beta \equiv V/c$, формула (107.6) переходит в

$$\omega' \approx \omega(1-\beta) = \omega(1 - V/c). \tag{107.7}$$

3. Рассмотрим теперь второй частный случай, когда наблюдение ведется в системе отсчета S' и притом перпендикулярно к направлению распространяющейся в ней волны. В этом случае удобнее воспользоваться формулой (107.4), полагая в ней $k'_{x'} = 0$. Это дает

$$\omega' = \omega\sqrt{1-\beta^2} \approx \omega(1 - \beta^2/2). \tag{107.8}$$

Происходит смещение частоты *в длинноволновую область спектра* во втором порядке по β. Это — чисто релятивистский эффект, невозможный в классической (дорелятивистской) теории. Он называется *поперечным эффектом Доплера*.

Для установления физического смысла поперечного эффекта Доплера предположим, что мимо наблюдателя прошел цуг из N волн. Наблюдатель измерил по своим часам время прохождения этого цуга и нашел для него значение t'. Очевидно, $\omega' = 2\pi N/t'$.

Найдем длительность того же процесса по часам системы S. Для этого надо отметить показания таких часов в моменты прохождения начала и конца цуга мимо наблюдателя. Разность этих показаний и определит длительность цуга t по часам системы S. Опять можно написать $\omega = 2\pi N/t$ и, следовательно, $\omega'/\omega = t/t'$. В этих измерениях время t' измеряется с помощью *одних и тех же часов*, неподвижных относительно наблюдателя, а для измерения времени t надо пользоваться *различными часами* системы S, синхронизованными между собой. Поэтому часы наблюдателя надо рассматривать как «неподвижные», а все часы системы S — как «движущиеся» (см. § 106, п. 2). Но движущиеся часы идут медленнее неподвижных, причем $t/t' = \sqrt{1-\beta^2}$, а потому $\omega' = \omega\sqrt{1-\beta^2}$. Таким образом, поперечный эффект Доплера есть не что иное, как проявление релятивистского эффекта *замедления хода движущихся часов*. Тем же эффектом объясняется появление квадратного корня $\sqrt{1-\beta^2}$ в формулах (107.2) и (107.4).

4. Поперечный эффект Доплера впервые удалось наблюдать Айвсу и Стилуэллу в 1938 г., а затем в 1941 г. Источником света служил пучок каналовых лучей атомов водорода, несущихся со значительной скоростью $v \sim 10^8$ см/с. Специальная конструкция трубки обеспечивала высокую однородность каналовых лучей по скоростям. Скорость v вычислялась по приложенному к электродам напряжению и по величине продольного доплеровского эффекта. Свет, испущенный движущимися атомами перпендикулярно к направлению их движения, попадал в спектрограф. По оценке авторов спектрограф давал возможность измерять длину волны с точностью 0,00025 нм. Главная трудность опыта состояла в том, что при отклонении направления пучка каналовых лучей от строгой перпендикулярности к линии наблюдения поперечное доплеровское смещение, пропорциональное $(v/c)^2$, могло оказаться перекрытым более сильным продольным эффектом Доплера, поскольку этот эффект — первого порядка по $(v\cos\vartheta)/c$ (ϑ — угол между скоростью атома и линией наблюдения). При скорости атома $v \approx 10^8$ см/с оба эффекта сравнимы между собой уже при $\vartheta \sim 5'$. Эта трудность была преодолена тем, что ось спектрографа устанавливалась перпендикулярно к плоскости зеркала. В спектрограф попадал не только свет, непосредственно излученный атомом, но и свет, отраженный от зеркала. Последний можно рассматривать как свет, излученный движущимся изображением атома. Если атом движется под углом ϑ к линии наблюдения, то его изображение в зеркале будет двигаться под углом $\pi - \vartheta$. Доплеровские смещения первого порядка для атома и его изображения в зеркале будут равны по величине, но противоположны по знаку. Если бы не было смещения второго порядка, то в спектрографе наблюдались бы две линии $\lambda_1 = \lambda - \Delta\lambda$ и $\lambda_2 = \lambda + \Delta\lambda$, симметрично расположенные относительно исходной линии λ. Эта симметрия будет нарушена поперечным доплеровским смещением $\delta\lambda$, которое происходит всегда в длинноволновую сторону спектра. С учетом этого $\lambda_1 = \lambda - \Delta\lambda + \delta\lambda$, $\lambda_2 = \lambda + \Delta\lambda + \delta\lambda$. Отсюда

$$\delta\lambda = \frac{\lambda_1 + \lambda_2}{2} - \lambda.$$

Таким путем удается отделить квадратичный (поперечный) эффект Доплера от линейного (продольного). Согласно (107.8), изменение длины волны при поперечном эффекте Доплера определяется выражением

$$\delta\lambda = \frac{1}{2}\lambda\beta^2. \tag{107.9}$$

В одном из опытов при напряжении на трубке 6788 В ожидаемое смещение, вычисленное по напряжению, должно было быть $\delta\lambda = 0{,}00116$ нм, а вычисленное по линейному доплеровскому смещению 0,00109 нм. Опыт дал $\delta\lambda = 0{,}0011$ нм. Подобное согласие получалось и при других напряжениях на трубке.

Результаты Айвса и Стилуэлла были подтверждены и другими учеными. Это были первые опыты, на которых впервые было экспериментально доказано *релятивистское замедление времени*, хотя сами Айвс и Стилуэлл, ничего не говоря о теории относительности, пытались интерпретировать их с точки зрения старой теории эфира.

5. Перейдем теперь к аберрации света. Введем единичные векторы $\mathbf{N}$ и $\mathbf{N}'$ в направлении распространения волны в системах отсчета S и S'. Тогда $\mathbf{k} = \omega\mathbf{N}/c$, $\mathbf{k}' = \omega'\mathbf{N}'/c$, и из формул (107.2) и (107.3) получим

$$N'_{x'} = \frac{N_x - \beta}{1 - \beta N_x}, \quad N'_{y'} = \frac{\sqrt{1-\beta^2}}{1 - \beta N_x}, \quad N'_{z'} = \frac{\sqrt{1-\beta^2}}{1 - \beta N_x}. \tag{107.10}$$

Аберрация света была открыта в 1727 г. королевским астрономом Брадлеем (1692–1762) при наблюдении звезд в телескоп. По этой причине ее называют также *астрономической аберрацией*. При описании этого явления в качестве «неподвижной» системы отсчета S возьмем систему Коперника (в которой Солнце считается неподвижным), а в качестве «движущейся» системы S' — систему, связанную с Землей. При этом мы полностью отвлечемся от вращения Земли вокруг собственной оси и будем рассуждать так, как если бы этого вращения совсем не было. Оси X и X' выберем в направлении орбитального движения Земли, скорость которого обозначим через V. Ось Y направим произвольно, но перпендикулярно к направлению движения Земли.

Обозначим далее через α угол между направлением на звезду (направлением оси телескопа) в системе S и положительным направлением оси Y (рис. 334). Тогда $N_x = -\sin\alpha$, $N_y = -\cos\alpha$, $N_z = 0$. Соответствующие величины в системе S' обозначим теми же буквами, но со штрихами. После подстановки в первую формулу (107.10) получим

$$\sin\alpha' = \frac{\sin\alpha + \beta}{1 + \beta\sin\alpha}, \qquad (107.11)$$

или

$$\sin\alpha = \frac{\sin\alpha' - \beta}{1 - \beta\sin\alpha'}. \qquad (107.12)$$

Формулами (107.11) и (107.12) и определяется направление распространения волны в одной из систем отсчета, если известно это направление в другой системе отсчета.

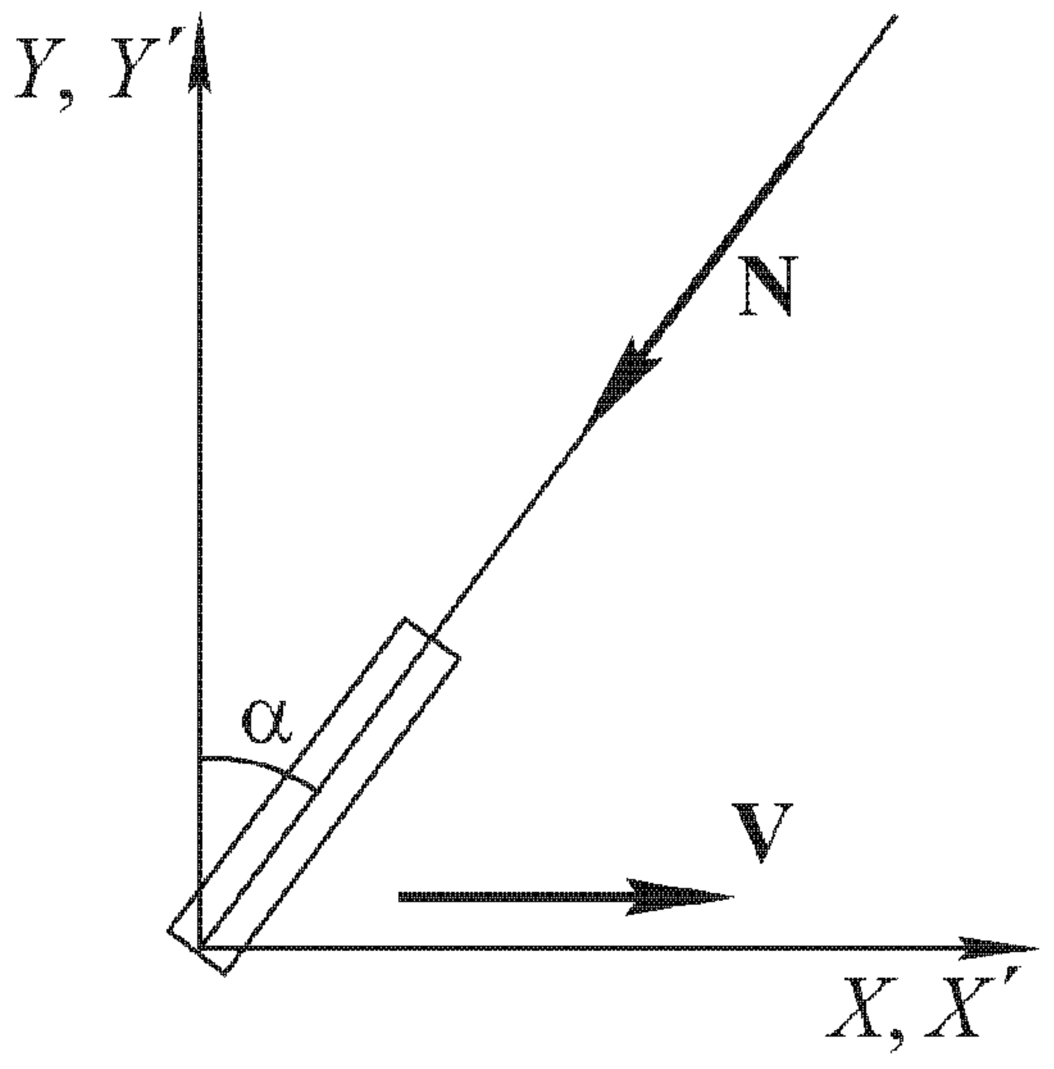


Рис. 334

Рассмотрим частный случай, когда свет в системе S распространяется вдоль оси Y. В этом случае $\alpha = 0$, и следовательно,

$$\sin\alpha' = \beta. \qquad (107.13)$$

Для того чтобы увидеть звезду в телескоп, его необходимо из перпендикулярного положения наклонить вперед на угол α', определяемый формулой (107.13). Угол α' называется *углом аберрации*. В аналогичном случае, когда свет в системе S' распространяется вдоль оси Y', т.е. когда $\alpha' = 0$, получим

$$\sin\alpha = -\beta. \qquad (107.14)$$

Телескоп надо наклонить на такой же угол, но уже назад. Скорость орбитального движения Земли $V \approx 30$ км/с, так что $\alpha \approx \approx V/c \approx 10^{-4}$ рад $\approx 20{,}5''$.

Если бы скорость $\mathbf{V}$ оставалась постоянной по величине и направлению, то видимое положение звезд на небесном своде оставалось бы неизменным и заметить аберрацию было бы нельзя. В действительности направление движения Земли по ее орбите непрерывно изменяется. Вследствие этого из-за аберрации будет изменяться и видимое положение звезд на небесном своде. Через полгода направление движения Земли изменится на противоположное, а угол аберрации изменит знак. Если звезда находится в плоскости земной орбиты, то ее видимое положение будет совершать в той же плоскости прямолинейное колебание с угловой амплитудой $\alpha = V/c = 20{,}5''$ и периодом в один год. Если направление на звезду (в системе S) перпендикулярно к плоскости земной орбиты, то видимое положение ее на небесном своде будет описывать окружность с угловым радиусом $20{,}5''$. Во всяком промежуточном положении траекторией видимого положения звезды будет эллипс, большая полуось которого имеет такие же угловые размеры. Так выглядело бы явление аберрации света в воображаемом идеальном случае, если бы единственным движением Земли было только ее движение по круговой орбите вокруг Солнца. В действительности аберрация света есть ничтожный эффект, накладывающийся на движение звезд по небесному своду, обусловленное вращением Земли вокруг собственной оси. Требовалась исключительная наблюдательность и экспериментальное искусство, чтобы выделить и измерить этот эффект. Это и было сделано Брадлеем.

Подчеркнем особо, что явление аберрации света не имеет никакого отношения к самой скорости движения звезд относительно Земли. Оно отражает только *изменение скорости* этого относительного движения, обусловленное движением Земли. Вот почему аберрация света одна и та же *для всех звезд*, хотя их скорости относительно Земли весьма различны.

6. Брадлей видел в явлении аберрации доказательство конечности скорости распространения света. Измерив угол аберрации α, он вычислил эту скорость но формуле $\alpha = V/c$. По времени это было второе (после Рёмера) измерение скорости света. Сам Брадлей истолковал открытое им явление с точки зрения ньютоновской корпускулярной теории света. В нерелятивистской кинематике скорости световой корпускулы $\mathbf{v}$ и $\mathbf{v}'$ в системах отсчета S и S' связаны соотношением $\mathbf{v} = \mathbf{v}' + \mathbf{V}$. Ограничимся случаем, когда в системе S световая корпускула движется по оси Y со скоростью $v_y = -c$. В системе S' появляется составляющая скорости корпускулы в перпендикулярном направлении $v'_x = v_x - V_x = -V$. Из-за этого траектория корпускулы в системе S' будет наклонена к оси Y под углом α', определяемым формулой

$$\operatorname{tg}\alpha' = \frac{v'_x}{v'_y} = \frac{V}{c} = \beta. \qquad (107.15)$$

Этой формулой в ньютоновской корпускулярной теории и определяется угол аберрации. При малых углах α' она совпадает с релятивистской формулой (107.13), если пренебречь квадратичными членами по β. Если бы к движению световой корпускулы применить релятивистский закон сложения скоростей (с учетом того, что скорость корпускулы по абсолютной величине равна c), то получились бы в точности прежние релятивистские формулы (107.10). Наблюдение аберрации света не позволяет, следовательно, сделать выбор между волновой и корпускулярной теориями света.

7. Истолкование аберрации в волновых теориях света, даже в их электромагнитной форме, встретило большие трудности, пока эти теории оставались механистическими. Все они были неудовлетворительны уже по той причине, что вводили произвольные предположения относительно механических свойств эфира и его движения относительно Земли. Впервые безупречная и в идейном отношении необычайно простая теория аберрации света (и оптического эффекта Доплера) была дана только в теории относительности.

В связи с этим затронем один вопрос, поставленный в эфирной теории света. Что будет с углом аберрации, если телескоп для его измерения заполнить водой? На этот вопрос эфирные теории света давали различный ответ, в зависимости от того, какие предположения вводили они о движении эфира относительно Земли и т.п. Опыт был поставлен Эйри в 1871 г. Оказалось, что при заполнении телескопа водой угол аберрации не изменяется. Объяснение этого результата в теории относительности не представляет затруднений. Для простоты рассуждений обратим направление распространения света, предположив, что источник света помещен в главном фокусе объектива телескопа. Поскольку нет никакого эфира, в системе отсчета, где телескоп покоится, вода или воздух, заполняющие его, а также стекло самого объектива телескопа оптически изотропны. В этой системе отсчета лучи выйдут из телескопа параллельно главной оптической оси, независимо от того, заполнен ли телескоп водой или не заполнен. Для определения угла аберрации надо выполнить переход к движущейся системе отсчета S'. Но это можно сделать для волны, *уже вышедшей из телескопа*. Направление этой волны совершенно не зависит от того, какой средой заполнен сам телескоп. Наличие телескопа и этой среды на таком переходе никак не отразится. Следовательно, и угол аберрации не будет зависеть от среды, заполняющей телескоп.

§ 108. Эффект Доплера в акустике и теории эфира

Эффект Доплера был открыт самим Доплером в 1842 г. на акустических волнах. В дальнейшем теория этого эффекта была перенесена без всяких изменений в оптику. При этом предполагалась справедливость волновой эфирной теории света. Место воздуха, в котором распространяются звуковые волны, в оптике занял световой эфир. В остальном все рассуждения в акустике и оптике были абсолютно тождественны. Эфирная теория безвозвратно ушла в область истории. Но акустический эффект Доплера полностью сохранил свое значение. Поэтому имеет смысл изложить здесь теорию этого эффекта. Поскольку, однако, этот том посвящен оптике, мы по-прежнему будем говорить о световых волнах в эфире. Для перехода к акустике слово «эфир» надо заменить словом «воздух», а световые волны — звуковыми.

В эфирной теории, помимо источника и наблюдателя, в явлении принимает участие промежуточная среда — световой эфир. С этим связаны усложнение и неопределенность эфирной теории эффекта Доплера, поскольку в каждом конкретном случае мы не можем сказать, как движутся источник и наблюдатель относительно «неуловимого» эфира. Различные теории эфира отличались друг от друга прежде всего тем, как они выбирали систему отсчета, в которой эфир покоится и, следовательно, ведет себя как оптически изотропная среда. Пусть в эфире распространяется плоская монохроматическая волна. Ее частоту в системе отсчета, в которой эфир покоится, обозначим через ω, а волновой вектор — через $\mathbf{k} = \omega\mathbf{N}/c$, где $\mathbf{N}$ — единичный вектор, в направлении которого волна распространяется. Фаза волны в указанной системе отсчета представится выражением $\varphi = \omega t - \mathbf{kr}$. Пусть относительно эфира равномерно движутся источник со скоростью $\mathbf{V}_\text{и}$ и наблюдатель со скоростью $\mathbf{V}_\text{н}$. Их радиусы-векторы будут соответственно $\mathbf{r}_\text{и} = \mathbf{V}_\text{и}t$ и $\mathbf{r}_\text{н} = \mathbf{V}_\text{н}t$. Фазы колебаний в этих движущихся точках определятся выражениями $\varphi_\text{и} = (\omega - \mathbf{kV}_\text{и})t$ и $\varphi_\text{н} = (\omega - \mathbf{kV}_\text{н})t$. Отсюда следует, что источник в системе отсчета, где он покоится, излучает волны с частотой $\omega_\text{и} = \omega - \mathbf{kV}_\text{и}$, а частота, воспринимаемая наблюдателем, определяется выражением $\omega_\text{н} = \omega - \mathbf{kV}_\text{н}$. Почленным делением с учетом соотношения $\mathbf{k} = \omega\mathbf{N}/c$ исключаем промежуточную частоту ω и находим

$$\frac{\omega_\text{н}}{\omega_\text{и}} = \frac{1 - (\mathbf{NV}_\text{н})/c}{1 - (\mathbf{NV}_\text{и})/c}. \tag{108.1}$$

Это и есть основная формула, определяющая доплеровское изменение частоты в теории эфира и в акустике. Мы видим, что в этой теории частота $\omega_\text{н}$ определяется движением как источника, так и наблюдателя *относительно эфира*, а также направлением распространения $\mathbf{N}$ волны в самом «неподвижном» эфире. В этом ее отличие от теории относительности, в которой эфира нет, а потому эффект Доплера зависит только от скорости источника *относительно наблюдателя* $(\mathbf{V}_\text{и} - \mathbf{V}_\text{н})$. В частности, при одной и той же относительной скорости $(\mathbf{V}_\text{и} - \mathbf{V}_\text{н})$ формула (108.1) приводит к разным результатам, в зависимости от того, что движется: источник или наблюдатель. Когда движется источник, а наблюдатель неподвижен, она дает

$$\omega_\text{н} = \frac{\omega_\text{и}}{1 - \mathbf{NV}_\text{и}/c}. \tag{108.2}$$

Если же движется наблюдатель, а источник остается в покое, то

$$\omega_\text{н} = \omega_\text{и}(1 - \mathbf{NV}_\text{н}/c). \tag{108.3}$$

В линейном приближении, когда в формуле (108.1) можно пренебречь квадратами обоих отношений $V_\text{и}/c$ и $V_\text{н}/c$, она переходит в

$$\frac{\omega_\text{н}}{\omega_\text{и}} = 1 + \frac{\mathbf{N}}{c}(\mathbf{V}_\text{и} - \mathbf{V}_\text{н}). \tag{108.4}$$

В эту формулу входит лишь относительная скорость $(\mathbf{V}_\text{и} - \mathbf{V}_\text{н})$, а не скорости $\mathbf{V}_\text{и}$ и $\mathbf{V}_\text{н}$ в отдельности. Однако формула еще не определяет изменения частоты, поскольку в нее входит также направление распространения волны $\mathbf{N}$ в «неподвижном» эфире. Действительно, хотя волна и посылается источником к наблюдателю, ее направление из-за аберрации будет изменяться с изменением движения эфира. Только в частном случае, когда источник или наблюдатель покоятся относительно эфира, эта неопределенность исчезает, а формула (108.4) совпадает с тем, что дает в первом порядке теория относительности. Тогда ей можно придать вид

$$\omega_\text{н} = (1 + V_N/c), \tag{108.5}$$

где V_N — скорость источника относительно наблюдателя по лучу зрения. Она считается положительной, когда источник приближается к наблюдателю, и отрицательной, когда он удаляется.

Все изложенное показывает, какие существенные упрощения и определенность внесла теория относительности в теорию эффекта Доплера и аберрации света.

ЗАДАЧА

На рис. 335 приведена схема опыта Фуко по определению скорости света в материальной среде. Лучи от источника S, пройдя через стеклянную пластинку M и линзу L, отражаются от плоского зеркала R, которое может вращаться вокруг оси, перпендикулярной к плоскости чертежа. Линза L дает изображение источника S на поверхности вогнутого зеркала C, центр кривизны которого совпадает с осью вращения зеркала R. Сосуд P наполняют исследуемым веществом, в котором измеряется скорость света. Если зеркало R неподвижно, то лучи, отраженные от C и R, снова сойдутся в точке S. Зеркало M отклонит часть лучей в сторону и даст действительное изображение источника в S_1. При вращении зеркала R изображение S_1 смещается в S_1'. По величине смещения S_1S_1' можно вычислить скорость света в исследуемом веществе. При отражении от движущегося зеркала происходит доплеровское изменение частоты света. Учитывая это, показать, что метод Фуко дает *групповую скорость света.*

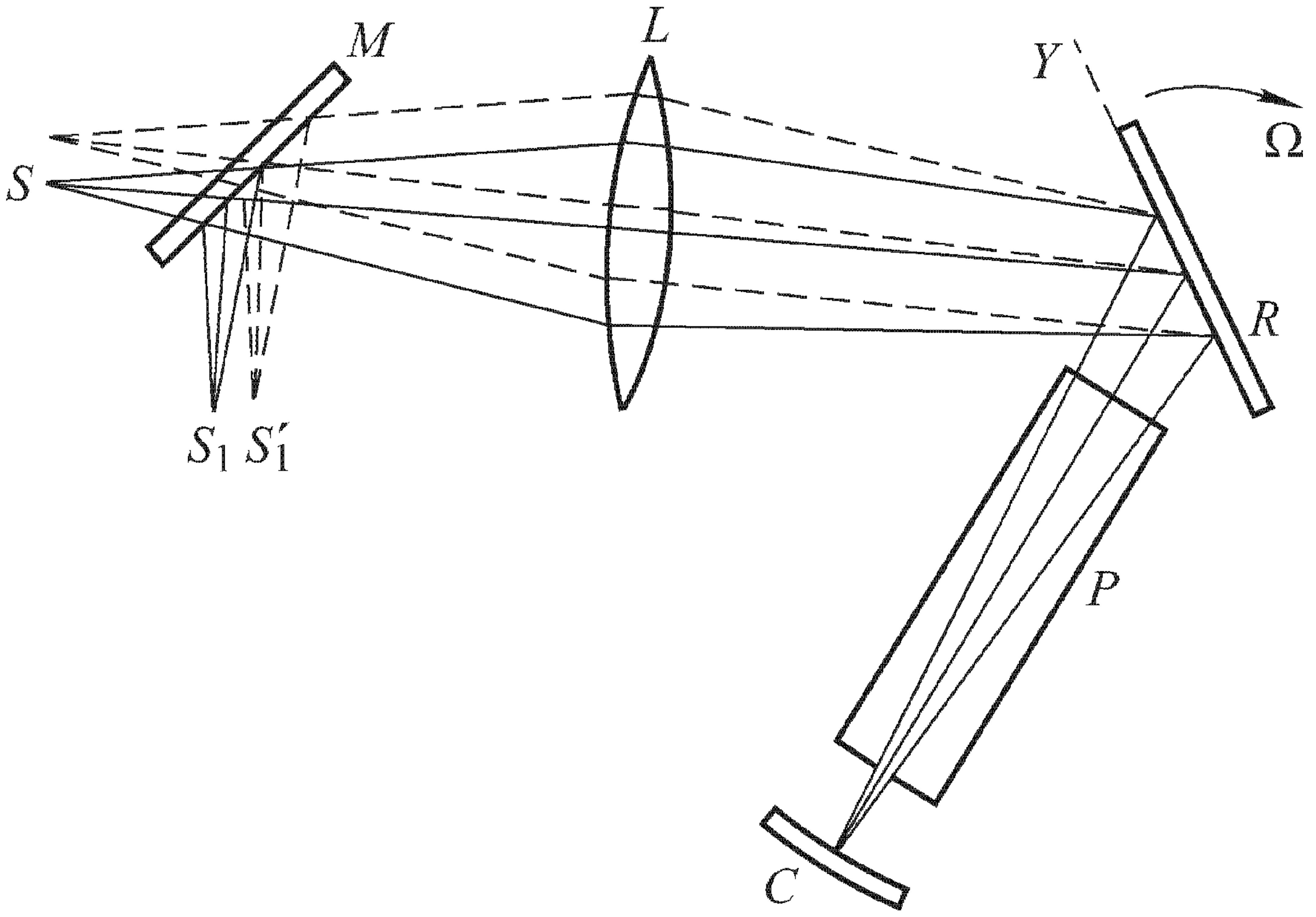


Рис. 335

Решение. Не теряя общности, воздушные зазоры между R и P, а также между P и C можно считать бесконечно тонкими и во всех расчетах пренебречь толщиной этих зазоров.

Обычный расчет проводится следующим образом. На прохождение расстояния от R до C и обратно волновой фронт, распространяющийся с фазовой скоростью v, затрачивает время $T = 2D/v$, где D — расстояние между зеркалами R и C. За это время R повернется на угол $\varphi = T\Omega$, если Ω — угловая скорость вращения зеркала. Луч, отраженный от зеркала R, вращается с вдвое большей скоростью. За то же время он повернется на угол

$\alpha = 2\varphi = 2T\Omega = (4D/v)\Omega$. Угол α легко рассчитать по величине смещения S_1S_1'. Таким образом,

$$v = 4D\Omega/\alpha. \tag{108.6}$$

В этом рассуждении не принято во внимание доплеровское изменение частоты при отражении света от вращающегося зеркала R. Поэтому оно не дает ответа, что за скорость вычисляется по формуле (108.6).

Поместим начало координат на оси вращающегося зеркала R и направим ось Y по линии пересечения плоскости этого зеркала с плоскостью чертежа. Так как линейные скорости различных точек вращающегося зеркала различны, то и изменение частоты волны при ее отражении от зеркала будет разным в зависимости от того, в каком месте зеркала произошло отражение. Благодаря этому различные точки волнового фронта будут распространяться в среде с различными фазовыми скоростями. Это поведет к вращению волнового фронта в среде. Если за направление положительного вращения принять направление вращения зеркала R, то для угловой скорости вращения волнового фронта в среде можно написать

$$\Omega' = \frac{1}{\cos\varphi}\frac{dv}{dy} = \frac{1}{\cos\varphi}\frac{dv}{d\omega}\frac{d\omega}{dy},$$

где φ — угол падения светового луча на зеркало R. Так как $v = \omega/k$, то

$$\frac{dv}{d\omega} = \frac{1}{k} - \frac{\omega}{k^2}\frac{dk}{d\omega} = \frac{v}{\omega} - \frac{v^2}{\omega u},$$

где u — групповая скорость. Остается определить $d\omega/dy$. Если ω — частота волны, отраженной от зеркала в точке с координатой y, а $\omega + d\omega$ — с координатой $y + dy$, то в первом порядке $d\omega/\omega = -\dfrac{2}{c}\,\Omega\cos\varphi\,dy$, откуда

$$\frac{1}{\cos\varphi}\frac{d\omega}{dy} = -2\Omega\,\frac{\omega}{c} = -2\Omega\,\frac{\omega}{nv},$$

где n — показатель преломления. Таким образом,

$$\Omega' = \left(\frac{v^2}{\omega u} - \frac{v}{u}\right)\cdot 2\Omega\frac{\omega}{nv} = \frac{2\Omega}{n}\left(\frac{v}{u} - 1\right).$$

Отраженный от зеркала C волновой фронт будет также поворачиваться при распространении в веществе с угловой скоростью Ω' и притом, как легко сообразить, в том же направлении, что и падающий волновой фронт. С другой стороны, на прохождение слоя вещества толщиной $2D$ волновой фронт затрачивает время $T = 2D/v$. За это время он повернется в среде на угол $\Omega'T = \dfrac{4D\Omega}{nv}\left(\dfrac{v}{u} - 1\right)$. По выходе из сосуда P в вакуум волновой фронт преломляется, вследствие чего угол поворота увеличивается в n раз и становится равным

$$n\Omega'T = \frac{4D\Omega}{v}\left(\frac{v}{u} - 1\right) = \frac{4D\Omega}{u} - \frac{4D\Omega}{v}.$$

Этот поворот надо прибавить к повороту $4D\Omega/v$, найденному ранее без учета эффекта Доплера. Таким образом, измеряемый угол поворота α в действительности равен

$$\alpha = \frac{4D\Omega}{v} + \left(\frac{4D\Omega}{u} - \frac{4D\Omega}{v}\right) = \frac{4D\Omega}{u},$$

так что вместо формулы (108.6) получится

$$u = 4D\Omega/\alpha. \tag{108.7}$$

Следовательно, метод вращающегося зеркала Фуко дает групповую скорость.

§ 109. Замедление хода часов в гравитационном поле

1. Вопросы релятивистской теории тяготения относятся к компетенции *общей теории относительности*, излагать которую мы не предполагаем. Затронем лишь кратко вопрос о влиянии поля тяготения на течение времени. Будем исходить из *принципа эквивалентности гравитационных сил и сил инерции*. Он был подробно изложен в § 71 первого тома. Гравитационное поле будем предполагать слабым (критерий слабости поля указывается ниже).

Возьмем какую-либо инерциальную систему отсчета S, в различных точках которой установлены неподвижные часы, синхронизованные по правилу Эйнштейна. Как всегда, время, определяемое по таким синхронизованным часам, будем обозначать через t. Пусть относительно S ускоренно движутся какие-то одни часы A. Время, отмечаемое часами A в точках, через которые они проходят, условимся обозначать через t_0. Часы A предполагаются идеальными, так что ускорение их само по себе не оказывает никакого влияния на физические процессы в часах. Такие часы идут одинаково быстро с часами в сопутствующих инерциальных системах отсчета. Промежутки времени по неподвижным и движущимся часам связаны соотношением

$$dt = dt_0/\sqrt{1-(v/c)^2}, \tag{109.1}$$

где v — скорость часов A относительно системы S. В инерциальной системе отсчета S нет никакого гравитационного поля. Единственной причиной замедления времени t_0 по сравнению с t может быть только движение часов A относительно инерциальной системы отсчета S.

2. Однако возможна и другая точка зрения. Допустим сначала, что часы A относительно системы S движутся с постоянным ускорением $\mathbf{a}$. Будем отсчитывать время t от того момента, когда скорость v была равна нулю. Тогда $v = \sqrt{2ax}$, где x — расстояние, которое прошли часы A за время t. Следовательно,

$$dt = dt_0/\sqrt{1-2ax/c^2}. \tag{109.2}$$

Теперь введем ускоренную систему отсчета S_0, движущуюся вместе с часами A. В этой системе часы A неподвижны, зато есть силы инерции. Если все явления описывать, приняв S_0 за систему отсчета, то причиной замедления времени t_0 следует считать *силы инерции*. Сила инерции, отнесенная к единице массы движущегося тела, равна $-\mathbf{a}$. Но, согласно принципу эквивалентности, силы инерции по своим физическим действиям неотличимы от гравитационного поля, напряженность которого в рассматриваемом нами случае равна $\mathbf{g} = -\mathbf{a}$. Введем еще *гравитационный потенциал* $\varphi = \pm gx$. Тогда предыдущая формула примет вид

$$dt = dt_0/\sqrt{1 - 2\varphi/c^2} \approx dt_0\left(1 - \frac{\varphi}{c^2}\right), \tag{109.3}$$

или

$$\frac{dt - dt_0}{dt_0} = -\frac{\varphi}{c^2}. \tag{109.4}$$

За нуль гравитационного потенциала принят потенциал такой точки, в которой движущиеся и неподвижные часы идут одинаково быстро. Поэтому в формулах (109.3) и (109.4) интервал времени dt можно отсчитывать не по часам инерциальной системы S, а по тем часам, покоящимся в системе S_0, которые находятся в точке B с нулевым потенциалом. Можно вообще начало отсчета гравитационного потенциала поместить в любой точке, если формулу (109.4) записать в виде

$$\frac{dt_{0A} - dt_{0B}}{dt_{0A}} = \frac{\varphi_B - \varphi_A}{c^2}, \tag{109.5}$$

где интервалы времени dt_{0A} и dt_{0B} отсчитываются по двум часам, покоящимся в ускоренной системе отсчета S_0 и помещенным в точках A и B с гравитационными потенциалами φ_A и φ_B. При этом на потенциалы должно быть наложено ограничение «слабого поля»:

$$\left|\frac{\varphi_B - \varphi_A}{c^2}\right| \ll 1, \tag{109.6}$$

так как только при этом условии приближенно справедлива ньютоновская теория тяготения и имеет смысл понятие гравитационного потенциала. (На поверхности Земли $\varphi/c^2 \approx -7\cdot 10^{-10}$, на поверхности Солнца $\varphi/c^2 \approx -2{,}12 \cdot 10^{-6}$, если за нуль принять потенциал в бесконечности.) Поэтому в формуле (109.5) безразлично, писать ли в знаменателе dt_{0A} или dt_{0B}.

Формула (109.5) имеет следующий смысл. Пусть в точках A и B гравитационного поля помещены тождественные часы, покоящиеся в системе отсчета S_0. Для сравнения темпа хода этих часов будем после каждого удара часов A посылать световые сигналы к часам B. Пусть N_A — число посланных сигналов. За время прихода сигналов часы в B совершат N_B ударов, причем

$$\frac{N_A - N_B}{N_A} = \frac{\varphi_B - \varphi_A}{c^2}. \tag{109.7}$$

Если $\varphi_A < \varphi_B$ (световой сигнал распространяется против направления гравитационного поля), то $N_A > N_B$, т.е. часы B идут медленнее часов A. В противоположном случае, когда $\varphi_B < \varphi_A$, медленнее будут идти часы A. Вообще, *часы в гравитационном поле идут тем медленнее, чем выше гравитационный потенциал в том месте, где они находятся.* Этот эффект, как и замедление хода движущихся часов, следует рассматривать как *свойство пространства-времени*, а не как влияние гравитационного поля на ход физических процессов в часах. Такое влияние, как уже указывалось в § 106, в случае реальных часов несомненно существует, но оно существенно зависит от того, какие физические явления положены в основу устройства часов. Для идеальных часов такого влияния нет — в этом случае эффект гравитационного замедления хода часов проявляется в чистом виде. Отмеченное обстоятельство проявляется и в формуле (109.5): гравитационное замедление времени есть *взаимное свойство двух часов*, определяющееся разностью гравитационных потенциалов между ними, тогда как для физических процессов в часах, очевидно, существен не потенциал, а напряженность гравитационного поля.

Конечно, формула (109.5) верна независимо от того, является ли гравитационное поле «фиктивным» полем сил инерции или «истинным» гравитационным полем, создаваемым массами вещества. Но эта формула была получена для однородного гравитационного поля. Применяя обычный прием, распространим ее на случай произвольного *неоднородного гравитационного поля.* Для этого вообразим, что в системе отсчета S_0 вдоль произвольной кривой AB, соединяющей часы A и B, расставлены достаточно часто тождественные часы $1, 2, 3, \ldots, n$. Гравитационное поле между каждыми соседними часами можно считать однородным и применить к ним формулу (109.5). Перемножив почленно все эти формулы и учтя, что в пределах их точности различием знаменателей можно пренебречь, мы снова придем к результату (109.5) и тем самым докажем его справедливость и для неоднородного поля.

Так как при наличии гравитационного поля темп хода даже тождественных часов в разных точках пространства разный, то теряет смысл тот способ синхронизации часов и установления одновременности пространственно разделенных событий, который применялся в инерциальных системах отсчета, т.е. при отсутствии гравитационного поля. При наличии гравитационного поля сохраняет смысл лишь *местное время*, устанавливаемое по тождественным часам в каждой точке пространства. Следует говорить не о синхронизации пространственно разделенных часов, а только о сравнении *темпа их хода*, как это описано выше.

В свете изложенного парадокс близнецов можно рассматривать с иной точки зрения. Часы брата A все время находятся в инерциальной системе отсчета, а часы брата B испытывают вместе с ним ускорение, что эквивалентно тому, что они неподвижны, но находятся в гравитационном поле. Это поле замедляет ход часов B. При возвращении к A часы B будут показывать меньшее время, чем A.

В последнее время (1976 г.) гравитационное замедление времени было подтверждено группой американских физиков Мерилендского университета. Измерялась разность показаний атомных часов на самолете и в наземной лаборатории. Самолет курсировал на высоте 10 км с небольшой скоростью, чтобы уменьшить кинематический эффект замедления времени. Время полета было около 15 часов. Ожидаемый гравитационный эффект составлял примерно +50 нс, кинематический −7 нс. Летающие часы сравнивались с наземными до, после и во время полета. Таким образом, можно было следить за монотонным возрастанием разности показаний сравниваемых часов. Опыт подтвердил теорию с точностью до 1,6 %.

3. Применим полученные результаты к излучению света. Пусть два тождественных неподвижных атома находятся в точках A и B гравитационного поля. Каждый из них в том месте, где он находится, излучает свет с одной и той же частотой ω_0. Рассматривая излучающие атомы как часы и применяя к ним формулу (109.5) или (109.7), получим

$$\frac{\omega_0 - \omega}{\omega_0} = \frac{\varphi_B - \varphi_A}{c^2}, \qquad (109.8)$$

где ω — частота света, приходящего от A в точку наблюдения B. Если свет распространяется от низшего потенциала φ_A к высшему φ_B (т.е. против гравитационного поля), то наблюдается изменение частоты *в красную сторону*. Для света, приходящего от поверхности Солнца к Земле, это изменение частоты составляет $(\omega_0 - \omega)/\omega_0 = 2{,}12 \cdot 10^{-6}$. Напротив, если свет распространяется от высшего гравитационного потенциала к низшему (т.е. в направлении гравитационного поля), то наблюдаемая частота света *увеличивается*. Это явление, называемое *гравитационным смещением спектральных линий*, было теоретически предсказано Эйнштейном, а затем подтверждено экспериментально. Несколько иначе описанное явление было рассмотрено в § 72 т. I.

§ 110. Сложение скоростей в теории относительности

1. Задача (физического) сложения скоростей формулируется следующим образом (см. т. I, § 7). Пусть скорость движущейся частицы в системе отсчета S' равна $\mathbf{v}'$, причем сама система S' равномерно движется относительно «неподвижной» системы

S со скоростью $\mathbf{V}$ в направлении оси X. Требуется определить скорость $\mathbf{v}$ той же частицы в системе отсчета S. Обозначим координаты частицы в системе S в момент времени t через x, y, z. Те же величины в системе S' обозначим штрихованными буквами t', x', y', z'. Так как обе эти группы переменных характеризуют *одно и то же событие* (прохождение частицы через одну и ту же пространственно-временную точку), то они связаны между собой формулами преобразования Лорентца (105.12). Поскольку $\mathbf{V} = \mathrm{const}$, координаты и время при движении частицы получают приращения

$$dx = \frac{dx' + V\,dt}{\sqrt{1-\beta^2}}, \quad dy = dy', \quad dz = dz', \quad dt = \frac{dt' + Vx'/c^2}{\sqrt{1-\beta^2}}.$$

Разделив первые три соотношения на последнее и замечая, что $dx/dt = v_x$, $dx'/dt' = v'_x$, получим

$$v_x = \frac{v'_x + V}{1 + v'_x V/c^2}, \quad v_y = \frac{v'_y\sqrt{1-\beta^2}}{1 + v'_x V/c^2}, \quad v_z = \frac{v'_z\sqrt{1-\beta^2}}{1 + v'_x V/c^2}. \qquad (110.1)$$

Эти формулы и выражают *правило сложения скоростей в релятивистской кинематике.* При медленных движениях, когда можно пренебречь квадратичными величинами V^2/c^2 и $v'_x V/c^2$, они переходят в нерелятивистские формулы

$$v_x = v'_x + V, \quad v_y = v'_y, \quad v_z = v'_z,$$

получающиеся из преобразования Галилея.

2. Для исследования формул (110.1) удобно принять за единицу скорость света c, введя безразмерные величины $\beta_x = v_x/c$, $\beta'_x = v'_x/c$, ... Тогда

$$\beta_x = \frac{\beta'_x + \beta}{1 + \beta\beta'_x}, \quad \beta_y = \frac{\beta'_y\sqrt{1-\beta^2}}{1 + \beta\beta'_x}, \quad \beta_z = \frac{\beta'_z\sqrt{1-\beta^2}}{1 + \beta\beta'_x}. \qquad (110.2)$$

Все скорости по величине не могут превосходить c, а соответствующие им безразмерные величины β, β'_x, ... — единицы. Возникает, однако, вопрос, не может ли в результате (физического) сложения двух скоростей получиться скорость, превосходящая c? Ответ на этот вопрос должен быть отрицательным.

Для доказательства рассмотрим сначала случай сложения параллельных скоростей, когда направления векторов $\mathbf{v}'$ и $\mathbf{V}$ совпадают или прямо противоположны. Результирующая скорость в этом случае полностью определяется первой формулой (110.1) или эквивалентной ей первой формулой (110.2). Так как β и β'_x не превосходят единицы, то их можно представить в виде $\beta = \operatorname{th}\vartheta$, $\beta'_x = \operatorname{th}\vartheta'_x$, где аргументы гиперболических тангенсов ϑ и ϑ'_x могут принимать любые вещественные значения. Тогда первая формула (110.2) запишется в виде

$$\beta_x = \frac{\operatorname{th}\vartheta'_x + \operatorname{th}\vartheta}{1 + \operatorname{th}\vartheta \operatorname{th}\vartheta'_x}.$$

Сравнивая ее с формулой сложения гиперболических тангенсов

$$\operatorname{th}(\vartheta'_x + \vartheta) = \frac{\operatorname{th}\vartheta'_x + \operatorname{th}\vartheta}{1 + \operatorname{th}\vartheta \operatorname{th}\vartheta'_x},$$

видим, что результирующую безразмерную скорость также можно представить гиперболическим тангенсом $\beta_x = \operatorname{th}\vartheta_x$, причем

$$\vartheta_x = \vartheta'_x + \vartheta. \tag{110.3}$$

Это значит, что (физическое) сложение параллельных скоростей сводится к алгебраическому сложению аргументов соответствующих им гиперболических тангенсов. Но при любых значениях аргументов гиперболический тангенс по абсолютной величине не может превосходить единицы. Следовательно, каковы бы ни были (параллельные) скорости составляющих движений, при их (физическом) сложении не может получиться скорость, превосходящая c. Если одна из складываемых скоростей равна c, то при сложении ее с любой другой скоростью получится также c. В этом и проявляется специфичность скорости света как предельной скорости, которую нельзя превзойти, сколько бы ни ускорять движение тела.

Для завершения доказательства остается исследовать случай, когда складываемые скорости взаимно перпендикулярны. Без ограничения общности можно считать, что скорость $\mathbf{v}'$ направлена вдоль оси Y'. Тогда, полагая в формуле (110.2) $\beta'_x = \beta'_z = 0$, получим

$$\beta_x = \beta, \quad \beta_y = \beta'_y\sqrt{1-\beta^2}, \quad \beta_z = 0.$$

Квадрат (безразмерной) скорости в системе S будет

$$\left(\frac{v}{c}\right)^2 = \beta_x^2 + \beta_y^2 = \beta^2 + \beta'^2_y(1-\beta^2).$$

Придав β постоянное, но произвольное значение, будем рассматривать величину $(v/c)^2$ как функцию аргумента β'^2_y. Производная этой величины $(1-\beta^2) \geqslant 0$, а потому функция $(v/c)^2$ возрастает с возрастанием β'^2_y. Она достигает максимума при $\beta'^2_y = 1$, который равен 1. При всех остальных значениях β'_y отношение $(v/c)^2$ не может превосходить единицы. Это и доказывает, что при сложении перпендикулярных скоростей также не может получиться скорость, превосходящая c.

3. Когда складываемые скорости параллельны, то индекс x можно опустить и написать

$$v = \frac{v' + V}{1 + v'V/c^2}. \tag{110.4}$$

Если $V \ll c$, то в этой формуле можно ограничиться первыми степенями по V. В этом приближении

$$v \approx (v' + V)\left(1 - \frac{v'V}{c^2}\right) \approx v' + \left(1 - \frac{v'^2}{c^2}\right)V. \tag{110.5}$$

Применим последнюю формулу к распространению света в жидкости, равномерно движущейся со скоростью V. Скорость света относительно неподвижной жидкости будет $v' = c/n$, где n — показатель преломления жидкости. Предполагая, что свет распространяется в направлении течения жидкости, найдем его скорость v относительно «неподвижной» системы отсчета. Для этого подставляем в формулу (110.5) $v'/c = 1/n$ и находим

$$v = \frac{c}{n} + \left(1 - \frac{1}{n^2}\right)V. \tag{110.6}$$

Формула (110.6) была получена еще Френелем в 1818 г. Он исходил из представления, что эфир увлекается движущимися телами, однако не полностью, а частично. Лоренц в электронной теории вновь пришел к формуле (110.6). Но истинный смысл ее был установлен только в теории относительности.

Формула (110.6) была экспериментально подтверждена Физо в 1851 г. Схема его опыта в усовершенствованном виде, какой ему придал Майкельсон в 1886 г., показана на рис. 336. Луч от источника S раздваивается разделительной пластинкой P. Один луч на рис. 336 изображен сплошной, а другой — штриховой линией. Затем лучи проходят через трубки, в которых течет вода. Один луч идет в направлении, а другой против течения воды. Из-за различия скоростей лучей относительно неподвижных стенок трубки между ними при выходе из прибора

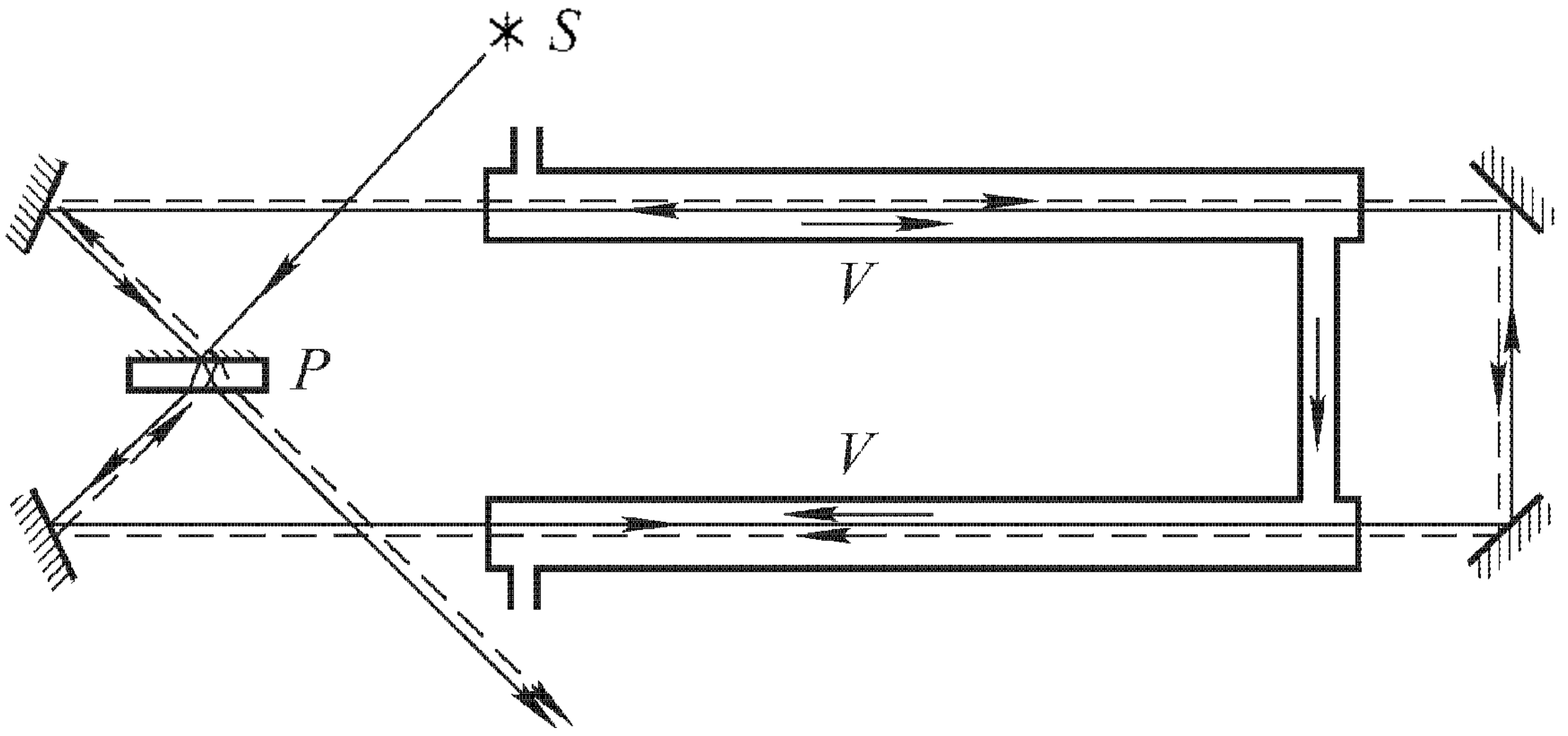


Рис. 336

возникает разность хода, изменяющаяся с изменением скорости течения V. Сначала наблюдается интерференция между лучами при неподвижной, а затем при текущей воде. По смещению интерференционных полос можно измерить разность хода, возникающую при течении, а по ней и разность скоростей $v - c/n$.

4. Лоренц в электронной теории обобщил формулу (110.6), учтя дисперсию света. Формулу Лоренца легко получить и в теории относительности. Надо только учесть доплеровское изменение длины волны, возникающее при течении воды. Обозначим через λ длину световой волны (в вакууме) в «неподвижной» системе отсчета S, а через $\lambda + \delta\lambda$ — в системе S', относительно которой вода неподвижна. Доплеровское изменение длины волны $\delta\lambda$ в первом порядке по V определяется выражением

$$\frac{\delta\lambda}{\lambda} = \frac{V}{v'} = \frac{Vn}{c}.$$

Следовательно,

$$v'(\lambda + \delta\lambda) = v'(\lambda) + \frac{dv'}{d\lambda}\frac{Vn}{c}\lambda,$$

или

$$v'(\lambda + \delta\lambda) = v'(\lambda) + \frac{d}{d\lambda}\left(\frac{c}{n}\right)\frac{Vn\lambda}{c} = \frac{c}{n(\lambda)} - \frac{\lambda}{n}\frac{dn}{d\lambda}V.$$

В формуле (110.5) под v' надо понимать, конечно, не $v'(\lambda)$, а $v'(\lambda + \delta\lambda)$. При этом второе слагаемое в правой части (110.5) в рассматриваемом приближении следует оставить без изменения. В результате получим

$$v = \frac{c}{n} + \left(1 - \frac{1}{n^2} - \frac{\lambda}{n}\frac{dn}{d\lambda}\right)V. \qquad (110.7)$$

Наличие добавочного члена, обусловленного дисперсией, было экспериментально подтверждено Зееманом в 1914 г.

Нелишне подчеркнуть, что доплеровское изменение длины волны относится к системе отсчета S', в которой вода неподвижна. Течение воды не сказывается на длине волны в «неподвижной» системе S, так что оба интерферирующих луча при выходе из прибора (см. рис. 336) имеют *одну и ту же длину волны.* Только благодаря этому и возможна интерференция.

§ 111. Релятивистская механика

1. Если какой-либо закон природы представлен в виде $A = B$, причем при переходе от одной системы отсчета к другой величины A и B остаются *неизменными*, то эти величины и самый закон называются *инвариантными* относительно этого перехода. Более общим является понятие *ковариантности.* Если при переходе от одной системы отсчета к другой величины A и B хотя и не остаются неизменными, но *преобразуются одинаково*, то закон $A = B$ сохраняется и в новой системе отсчета. В этом случае говорят, что закон $A = B$ *ковариантен* относительно рассматриваемого преобразования систем отсчета. Часто термин «инвариантность закона» употребляют в смысле его ковариантности.

До теории относительности допустимыми считались только *галилеевы преобразования координат.* Относительно этих преобразований уравнения механики Ньютона были ковариантны (инвариантны), тогда как уравнения электродинамики Максвелла–Лорентца — не ковариантны. Теория относительности показала, что от галилеева преобразования надо отказаться и заменить его *преобразованием Лорентца.* Тогда принцип относительности требует, чтобы *законы природы были ковариантны относительно преобразования Лорентца.* Этому требованию уравнения электродинамики удовлетворяют, а уравнения механики Ньютона не удовлетворяют. Поэтому механика Ньютона должна быть изменена.

В ньютоновской механике сила, действующая на тело в какой-то момент времени, определяется положением всех взаимодействующих тел *в тот же момент.* Но в теории относительности понятие «тот же момент времени» зависит от выбора системы отсчета. Невозможно автоматически преобразовать каждый закон сил ньютоновской механики в лорентц-ковариантную форму. Допустимы только такие теории, из которых *может быть исключено понятие действия на расстоянии.* Такая возможность существует в *теории столкновений.* Последняя исходит из идеализированного представления, что взаимодействие имеет место только в продолжение того промежутка времени, когда расстояние между телами или точечными частицами бесконечно мало по сравнению с размерами самих тел или другими характерными расстояниями, определяющими характер процессов столкновения. До и после этого бесконечно малого промежутка времени тела движутся свободно. К процессам столкновений применимы законы сохранения импульса и энергии, но им надо придать лорентц-ковариантную форму. Это и является целью настоящего параграфа. Дальнодействие можно исключить также при рассмотрении движения электрически заряженных частиц в электромагнитных полях. Однако изложение относящихся сюда вопросов электродинамики потребовало бы слишком много места, а потому мы ограничимся только рассмотрением процессов столкновения.

2. Для решения поставленной задачи проще и логичнее всего воспользоваться понятием *четырехмерного вектора* (или, короче, *4-вектора*) в пространстве Минковского. Каждое «точечное» событие в таком пространстве характеризуется совокупностью четырех координат x, y, z, $\tau \equiv ct$. При переходе от системы отсчета S к системе отсчета S' разности координат двух точек преобразуются по формулам

$$\Delta x' = \frac{\Delta x - \beta\,\Delta\tau}{\sqrt{1-\beta^2}}, \quad \Delta y' = \Delta y, \quad \Delta z' = \Delta z, \quad \Delta\tau' = \frac{\Delta\tau - \beta\,\Delta x}{\sqrt{1-\beta^2}}, \tag{111.1}$$

как это следует из (105.12). Напомним, что квадрат интервала между рассматриваемыми точками есть инвариант:

$$\Delta\tau^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 = \mathrm{Inv}\ . \tag{111.2}$$

Мы воспользовались частным преобразованием Лорентца (105.12), в котором предполагается, что координатные оси X, Y, Z параллельны осям X', Y', Z', а система S' движется относительно S вдоль оси X. Можно было бы взять любую ориентацию осей и любое направление движения, но это только усложнило бы запись, ничего не меняя по существу.

Назовем четырехмерным вектором совокупность четырех величин A_x, A_y, A_z, A_τ, которые при переходе от одной системы отсчета к другой преобразуются так же, как разности координат двух точек в пространстве Минковского, т.е.

$$A'_{x'} = \frac{A_x - \beta A_\tau}{\sqrt{1-\beta^2}}, \quad A'_{y'} = A_y, \quad A'_{z'} = A_z, \quad A'_{\tau'} = \frac{A_\tau - \beta A_x}{\sqrt{1-\beta^2}}. \tag{111.3}$$

Величины A_x, A_y, A_z называются *пространственными*, а A_τ — *временной* составляющей четырехмерного вектора. Пространственные составляющие мы объединим в обычный трехмерный вектор $\mathbf{A}$ и будем обозначать четырехмерный вектор через $(\mathbf{A}, A_\tau)$. Из тождественности законов преобразования (111.1) и (111.3) следует, что четырехмерный вектор $(\mathbf{A}, A_\tau)$ обладает *инвариантом*:

$$A_\tau^2 - \mathbf{A}^2 = \mathrm{Inv}\ . \tag{111.4}$$

Если какой-либо закон природы записан в четырехмерной векторной форме $(\mathbf{A}, A_\tau) = (\mathbf{B}, B_\tau)$, то он лоренцковариантен, так как обе части написанного равенства при преобразовании Лорентца преобразуются одинаково. Четырехмерная векторная форма эквивалентна двум уравнениям: $\mathbf{A} = \mathbf{B}$ и $A_\tau = B_\tau$. Этим замечанием мы и воспользуемся для решения поставленной задачи. Именно, мы постулируем, что *закон сохранения импульса и энергии можно записать в виде равенства четырехмерных векторов.* Задача состоит в том, чтобы найти вид этих векторов.

3. Рассмотрим частицу, движущуюся со скоростью $\mathbf{v}$ относительно неподвижной системы отсчета S. Пусть $d\mathbf{r}$ — ее перемещение за время $dt = d\tau/c$. Эти величины образуют четырехмерный вектор $(d\mathbf{r}, c\,dt)$. Очевидно, он останется четырехмерным вектором и после умножения его составляющих на одну и ту же постоянную. Возьмем в качестве таковой m_0/dt_0, где m_0 — некоторая постоянная, а $dt_0 = dt\sqrt{1 - v^2/c^2}$ — собственное время, которое, как известно, является инвариантом. Тогда получим четырехмерный вектор

$$m_0\left(\frac{d\mathbf{r}}{dt_0}, c\,\frac{dt}{dt_0}\right) = m(\mathbf{v}, c), \tag{111.5}$$

где введено обозначение

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}. \tag{111.6}$$

Допустим теперь, что частица движется *медленно*, так что величиной v^2/c^2 можно пренебречь. Возьмем в качестве m_0 массу частицы, как она определяется в нерелятивистской механике. Тогда пространственная составляющая четырехмерного вектора (111.5) будет $m_0\mathbf{v}$. Вектор $m_0\mathbf{v}$ в ньютоновской механике называется *импульсом*. Поэтому в релятивистской механике естественно определить импульс выражением

$$\mathbf{P} = m\mathbf{v} = \frac{m_0\mathbf{v}}{\sqrt{1 - v^2/c^2}}, \tag{111.7}$$

поскольку в пределе при малых скоростях оно переходит в нерелятивистское выражение $m_0\mathbf{v}$. Величина m_0 называется *массой покоя*, а m — *массой движения* или *релятивистской массой*. Таким образом, $(\mathbf{P}, mc)$ есть четырехмерный вектор, а величина $(mc)^2 - P^2$ — его инвариант. Значение этого инварианта легко найти: при $v = 0$ он обращается в $(m_0c)^2$, а потому

$$(mc)^2 - P^2 = (m_0c)^2. \tag{111.8}$$

Остается выяснить физический смысл временной части четырехмерного вектора (111.5). Для этого замечаем, что $d\mathbf{P}/dt$ есть сила, действующая на частицу. Работа этой силы на перемещении $\mathbf{v}\,dt$ равна $dA = \mathbf{v}\,d\mathbf{P} = \mathbf{P}\,d\mathbf{P}/m$ или на основании (111.8) $dA = c^2\,dm$. Энергия частицы найдется интегрированием этого выражения по m. Если постоянную интегрирования положить равной нулю, то получится формула

$$\mathcal{E} = mc^2 = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}}. \tag{111.9}$$

Величина $\mathcal{E}$ называется *полной энергией частицы*. Для покоящейся частицы $m = m_0$, так что (111.9) переходит в

$$\mathcal{E}_0 = m_0c^2. \tag{111.10}$$

Величина $\mathcal{E}_0$ называется *энергией покоя частицы*.

Формула (111.9), впервые в общем виде полученная Эйнштейном, устанавливает *взаимосвязь между массой и энергией*. Кинетическая энергия частицы определяется выражением

$$K = \mathcal{E} - \mathcal{E}_0 = (m - m_0)c^2. \tag{111.11}$$

При медленных движениях это выражение переходит в обычную формулу $K = {}^1\!/_2 m_0v^2$.

Импульс и энергия теперь объединены в четырехмерный вектор

$$(\mathbf{P}, mc) = (\mathbf{P}, \mathcal{E}/c), \tag{111.12}$$

называемый *вектором импульса–энергии*. Его инвариантом относительно преобразования Лоренца является

$$\left(\frac{\mathcal{E}}{c}\right)^2 - P^2 = (m_0 c)^2 = \mathrm{Inv}\ . \tag{111.13}$$

Тем самым в теории относительности законы сохранения импульса и энергии перестают быть независимыми законами, а объединяются в *единый закон сохранения четырехмерного вектора импульса–энергии*. Его называют также *законом сохранения импульса–энергии*.

Остается ответить на два вопроса. Во-первых, почему при вычислении работы сила была определена так же, как в нерелятивистской механике, т.е. как производная $d\mathbf{P}/dt$? Во-вторых, почему энергия всегда определяется с точностью до несущественной произвольной постоянной, здесь же она определена *однозначно*? Ответ на оба вопроса, в сущности, один и тот же. Он состоит в том, что на величину $\mathcal{E}$, вычисленную выше и названную полной энергией, было наложено требование, чтобы она (после деления на c) была *временной компонентой четырехмерного вектора* (111.5). Если энергию не определить однозначно, то она этому требованию удовлетворять не будет.

Для системы невзаимодействующих частиц, а также частиц, взаимодействующих только при столкновениях, четырехмерный вектор импульса–энергии определяется как сумма четырехмерных векторов импульса–энергии этих частиц. При этом в теории относительности достигается однообразная трактовка упругих и неупругих столкновений. Независимо от характера столкновения сохраняется трехмерный вектор импульса системы. Следовательно, должна сохраняться и энергия, как (умноженная на c) временная компонента четырехмерного вектора. Вместе с энергией сохраняется и релятивистская масса. Только при упругих и неупругих столкновениях она по-разному распределяется между массой покоя и массой, связанной с кинетической энергией макроскопического движения. Например, при столкновении двух одинаковых неупругих шаров, движущихся с одинаковыми скоростями навстречу друг другу, исчезновение кинетической энергии макроскопического движения (т.е. переход его во внутреннее молекулярное движение) проявляется в эквивалентном увеличении массы покоя системы: масса нагретого шара больше, чем масса такого же холодного шара. При упругих же столкновениях остаются неизменными и масса покоя, в масса, связанная с кинетической энергией макроскопического движения.

Существуют частицы (*фотоны, нейтрино*), для которых масса покоя равна нулю. Для них связь (111.13) между энергией и импульсом имеет вид

$$P = \mathcal{E}/c. \tag{111.14}$$

Такие частицы всегда движутся со скоростью c. Иначе, как видно из формул (111.7) и (111.9), импульс и энергия таких частиц обращались бы в нуль.

ЗАДАЧИ

1. Две одинаковые частицы движутся в лабораторной системе навстречу друг другу с одной и той же скоростью v. Найти относительную скорость V каждой из них относительно другой. Какой энергией $\mathcal{E}'$ в лабораторной системе отсчета должна обладать одна из частиц, чтобы получить ту же относительную скорость, если вторая частица (мишень) неподвижна? (Принцип действия *ускорителя на встречных пучках.*)

Решение. По теореме сложения скоростей

$$V = \frac{2v}{1 + v^2/c^2}.$$

Искомая полная энергия, которую надо было бы сообщить одной частице, равна

$$\mathcal{E}' = \frac{\mathcal{E}_0}{\sqrt{1 - V^2/c^2}},$$

где $\mathcal{E}_0$ — энергия покоя частицы. Фактическая энергия, которой обладает частица, $\mathcal{E} = \mathcal{E}_0/\sqrt{1 - v^2/c^2}$. Отсюда нетрудно получить

$$\mathcal{E}' = 2\frac{\mathcal{E}^2}{\mathcal{E}_0} - \mathcal{E}_0, \tag{111.15}$$

а для кинетической энергии

$$K' = 2\left(\frac{\mathcal{E}^2}{\mathcal{E}_0} - \mathcal{E}_0\right). \tag{111.16}$$

(Другое решение см. в т. I, § 28.)

2. Вывести формулу, являющуюся релятивистским обобщением формулы Циолковского (см. т. I, § 21) для движения ракеты. Считать, что скорости ракеты и газовой струи направлены вдоль одной прямой.

Решение. На основании законов сохранения импульса и энергии

$$mv + m_{\text{газ}}v_{\text{газ}} = \text{const}, \quad m + m_{\text{газ}} = \text{const},$$

где m и $m_{\text{газ}}$ — релятивистские массы ракеты и газов, а v и $v_{\text{газ}}$ — их скорости в произвольный момент времени. Газы, уже покинувшие ракету, не влияют на ее движение. Поэтому можно считать, что в рассматриваемый момент времени $m_{\text{газ}} = 0$. Тогда не возникает неопределенности, что следует понимать под $v_{\text{газ}}$. Однако, поскольку газы непрерывно образуются, $dm_{\text{газ}} \neq 0$. Дифференцируя предыдущие уравнения, получим

$$m\,dv + (v - v_{\text{газ}})\,dm = 0.$$

По релятивистскому закону сложения скоростей

$$v_{\text{газ}} = \frac{v-u}{1-vu/c^2},$$

где u — скорость газовой струи относительно ракеты. Исключение $v_{\text{газ}}$ приводит к уравнению

$$dv + u\frac{1-v^2/c^2}{1-vu/c^2}\frac{dm}{m} = 0.$$

Воспользовавшись формулой (111.6), после несложных преобразований найдем

$$\frac{dv}{v^2-c^2} = \frac{u}{c^2}\frac{dm_0}{m_0}.$$

Предполагая скорость u газовой струи постоянной и интегрируя, получим искомый результат:

$$\frac{(m_0)_{\text{нач}}}{m_0} = \left(\frac{1+v/c}{1-v/c}\right)^{c/2u}. \qquad (111.17)$$

ГЛАВА X

ТЕПЛОВОЕ ИЗЛУЧЕНИЕ

§ 112. Равновесное излучение в полости

1. Излучение света происходит в результате переходов атомов, молекул и других атомных систем из состояний с большей в состояния с меньшей энергией. Так называемое *тепловое* или *температурное излучение* отличается от других видов излучения (люминесценции) только способом перехода излучающих систем в возбужденные состояния. В явлениях теплового излучения такой переход осуществляется в результате теплового движения атомов и молекул.

В проблемах теплового излучения особо важное значение имеет понятие так называемого *равновесного излучения*. Для установления этого понятия рассмотрим полость с неподвижными и непрозрачными стенками, температура которых поддерживается постоянной. Атомы и молекулы стенок переходят в возбужденные состояния за счет энергии теплового движения и при обратных переходах в невозбужденные состояния дают излучение, заполняющее полость. Падая на стенки полости, лучистая энергия частично отражается, частично поглощается. Происходит изменение направления распространения, спектрального состава, поляризации, интенсивности излучения. В результате всех этих процессов, как это следует из общего начала термодинамики, в полости в конце концов устанавливается макроскопически вполне определенное состояние излучения, при котором за каждый промежуток времени количество излученной лучистой энергии определенного цвета, направления распространения и поляризации в среднем равно количеству поглощенной энергии того же цвета, направления распространения и поляризации. Как и всякое равновесное состояние, оно характеризуется тем, что каждому микропроцессу, происходящему в системе, с той же вероятностью соответствует микропроцесс, идущий в обратном направлении (*принцип детального равновесия*). Благодаря этому состояние излучения в полости и остается *макроскопически неизменным* во времени. Переход в равновесное состояние, как и всякий статистический процесс, управляется *вероятностными законами.* В полости устанавливается хаотическое состояние излучения, которому соответствует *наибольшая вероятность.* Оно и называется равновесным излучением.

Свойства равновесного излучения: плотность лучистой энергии, ее распределение по спектру частот и направлениям распространения, а также поляризация излучения *совершенно не зависят от формы и материала стенок полости.* Эти свойства, подобно состоянию газа в сосуде, определяются *только температурой стенок полости.* Равновесное излучение *однородно*, т.е. его плотность одна и та же во всех точках внутри полости. Оно *изотропно и неполяризовано*: все возможные направления распространения излучения представлены с одинаковой вероятностью, а направления векторов **E** и **B** и каждой точке пространства хаотически меняются во времени. Поскольку излучение находится в тепловом равновесии со стенками, можно говорить о температуре не только стенок, но и о *температуре самого излучения*, считая по определению обе температуры равными. Надо, однако, подчеркнуть, что температура равновесного излучения есть *свойство самого излучения*, а не стенки, с которой оно находится в тепловом равновесии. О ней имеет смысл говорить и тогда, когда вообще нет никакой стенки. В частности, например, плотность энергии равновесного излучения однозначно определяет и его температуру (см. § 115).

Если стенки полости совершенно непоглощающие, например идеально зеркальные, то в такой полости не будет поглощения и испускания света. В полость можно ввести излучение произвольного спектрального состава. Отражаясь от стенок, излучение меняет направление распространения, но его спектральный состав сохраняется неизменным. При надлежащей геометрической форме полости с зеркальными стенками возможны и такие случаи, когда сохраняются также направление распространения и поляризация излучения. Так будет, например, когда полость имеет форму прямого цилиндра с абсолютно зеркальными основаниями. Тогда луч света произвольной частоты и поляризации может распространяться туда и обратно параллельно оси цилиндра, последовательно отражаясь от зеркальных оснований. Но все подобные случаи являются идеальными и никогда точно не реализуются в действительности. Излучение в полости в этих случаях *неравновесно* и *неустойчиво.* Уже при сколь угодно малых отклонениях от идеальности, если только подождать достаточно долго, в полости обязательно установится равновесное излучение. Идеальные системы, однако, имеют большое значение в теоретических рассуждениях. Можно брать стенки абсолютно зеркальными и в то же время считать, что в полости всегда устанавливается равновесное излучение. Для этого достаточно, например, ввести в полость сколь угодно малое поглощающее и излучающее тельце — *пылинку*, по выражению Планка. Такая пылинка, практически не играя никакой роли в энергетическом балансе системы, переводит, однако, любое неравновесное состояние, возникшее в полости, в равновесное.

2. Введем теперь некоторые величины, характеризующие состояние излучения в пространстве. Эти величины имеют смысл для произвольного излучения, а не только для равновесного.

Обозначим через u плотность энергии излучения, т.е. количество такой энергии в единице объема пространства. Ее можно разложить по частотам или длинам волн, т.е. представить в виде

$$u = \int_0^\infty u_\omega \, d\omega = \int_0^\infty u_\lambda \, d\lambda. \tag{112.1}$$

Величины $u_\omega \, d\omega$ и $u_\lambda \, d\lambda$ имеют смысл объемной плотности лучистой энергии, приходящейся на интервал частот ω, $\omega + d\omega$ или интервал длин волн λ, $\lambda + d\lambda$. Коэффициенты u_ω и u_λ называются *спектральными плотностями лучистой энергии*. Если речь идет об одном и том же спектральном интервале, но представленном в различных формах, то $u_\omega \, d\omega = u_\lambda \, d\lambda$. При этом $\lambda = 2\pi c/\omega$, и следовательно, $d\lambda/\lambda = -d\omega/\omega$. Знак минус мы опустим — он означает только, что с возрастанием частоты длина волны убывает. Таким образом, считая величины $d\omega$ и $d\lambda$ существенно положительными, можно написать

$$u_\lambda = \frac{\omega}{\lambda} u_\omega, \qquad u_\omega = \frac{\lambda}{\omega} u_\lambda. \tag{112.2}$$

В теоретической физике обычно пользуются величиной u_ω, в экспериментальной отдают предпочтение u_λ. В случае равновесного излучения функции u_ω и u_λ зависят только от частоты ω (или длины волны λ) и от температуры излучения T, но не зависят от формы и материала стенок полости. Они зависят от среды, занимающей полость. Эта зависимость будет установлена в § 114. Пока же будем предполагать, что в полости вакуум. В этом случае u_ω будет *универсальной функцией только* ω и T, а u — *универсальной функцией только* T. Нахождение функции $u_\omega(\omega, T)$ является *основной задачей* теории теплового излучения.

Поток лучистой энергии, проходящей за время dt через малую площадку ds в пределах телесного угла $d\Omega$, ось которого перпендикулярна к площадке ds, можно представить выражением

$$d\Phi = I \, ds \, d\Omega \, dt. \tag{112.3}$$

Величина I называется *удельной интенсивностью излучения*. Если ее разложить в спектр, т.е. представить интегралом

$$I = \int\limits_0^\infty I_\omega\, d\omega, \tag{112.4}$$

то величину I_ω называют *удельной интенсивностью излучения частоты* ω.

Найдем связь между u и I (а также между u_ω и I_ω) для поля излучения в вакууме. Возьмем в пространстве бесконечно малый прямоугольный параллелепипед с площадью основания ds и высотой dl (рис. 337). Выделим пучок лучей, вступающих через площадку ds внутрь параллелепипеда, направления которых лежат в пределах телесного угла $d\Omega$, а ось телесного угла нормальна к основанию ds. Каждый из этих лучей доходит до второго основания параллелепипеда за время $dt = dl/c$. За это время из выделенного пучка лучей через площадку ds внутрь параллелепипеда вступает лучистая энергия $I\,ds\,d\Omega\,dt = (I/c)\,d\Omega\,dV$, где $dV = dl\,ds$ — объем параллелепипеда. Разделив на dV, найдем $(I/c)\,d\Omega$. Эта величина есть плотность du лучистой энергии, распространяющейся в пределах телесного угла $d\Omega$. По своему смыслу она может быть функцией точки в пространстве, но не может зависеть от формы параллелепипеда. Поэтому для нахождения полной плотности u лучистой энергии в рассматриваемой точке пространства надо проинтегрировать выражение du по всем направлениям в пространстве. В общем случае при таком интегрировании надо принять во внимание, что удельная интенсивность I зависит от направления излучения. Но в случае изотропного излучения, каким, в частности, является равновесное излучение, такой зависимости нет. В этом случае интегрирование сводится к простой замене элементарного телесного угла $d\Omega$ на полный телесный угол $\Omega = 4\pi$. Тогда получится

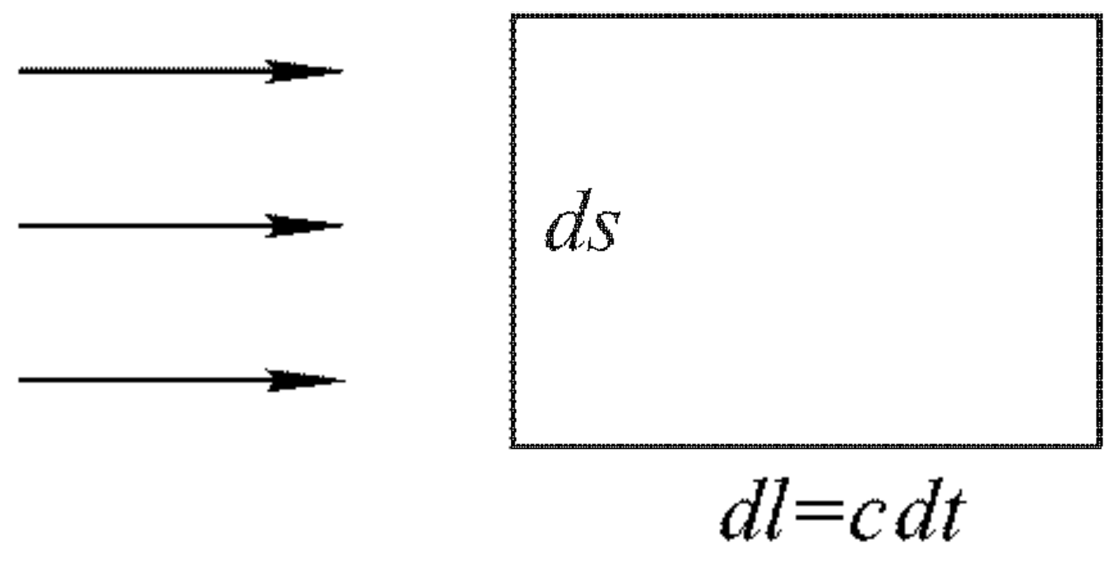


Рис. 337

$$u = \frac{4\pi}{c} I. \tag{112.5}$$

Аналогично,

$$u_\omega = \frac{4\pi}{c} I_\omega. \tag{112.6}$$

Такие же соотношения справедливы и в том случае, когда полость заполнена однородной изотропной непоглощающей средой. Надо только величину c заменить групповой скоростью света в рассматриваемой среде.

§ 113. Закон Кирхгофа

1. Перейдем теперь к рассмотрению *законов теплового излучения и поглощения тел.* Ограничимся частным случаем, когда излучающее и поглощающее тело *непрозрачно.* Этому условию можно удовлетворить, если толщина тела достаточна, чтобы всякий луч, вступивший в него, успел поглотиться, не достигнув противоположной границы тела. Поток лучистой энергии с частотами между ω и $\omega + d\omega$, излучаемый за время dt площадкой ds поверхности тела внутрь телесного угла $d\Omega$, можно представить выражением

$$d\Phi = E_\omega \, ds \cos\varphi \, d\Omega \, d\omega \, dt, \tag{113.1}$$

где φ — угол между направлением излучения и нормалью к площадке ds. Здесь введен множитель $\cos\varphi$, так как излучение удобнее относить не к величине полной площади ds, а к ее «видимой» части, т.е. проекции площадки ds на плоскость, перпендикулярную к направлению излучения (см. § 22). Величина E_ω называется *излучательной способностью* поверхности тела в направлении, определяемом углом φ [1].

Поглощательной способностью A_ω поверхности для излучения той же частоты, направления распространения и поляризации называется безразмерная величина, показывающая, какая доля энергии падающего излучения поглощается рассматриваемой поверхностью.

Речь идет, конечно, только о *чисто температурном излучении*, когда все свойства излучающего и поглощающего тела определяются *только температурой тела.* Исключаются из рассмотрения все случаи «холодного» или неравновесного свечения, когда светящееся тело излучает не потому, что оно нагрето, а потому, что определенные уровни энергии его возбуждены и тело постепенно «высвечивается», переходя в нормальное, невозбужденное состояние.

Что касается зависимости излучательной способности тела от окружающей однородной среды, то к этому вопросу можно подходить с двух точек зрения. Можно определить E_ω через *полный поток* лучистой энергии выделенного направления и частоты, исходящий от поверхности тела в *окружающую среду.* При такой точке зрения излучательная способность характеризует свойства *двух сред*, граничащих между собой вдоль рассматриваемой поверхности. Но можно встать на другую точку зрения, развивавшуюся Прево (1751–1839) в начале прошлого века. Согласно этой точке зрения, результирующий поток лучистой энергии через поверхность тела есть *разность двух потоков*, которые оба излучаются в вакуум: потока, *излучаемого телом* в окружающее пространство, и потока, *излучаемого средой* в пространство, занятое телом. Первый поток определяет излучательную способность *тела*, второй — излучательную способность *среды*. Мы будем придерживаться второй точки зрения, так как тогда величины E_ω и A_ω будут характеристиками только *самого тела*, не зависящими от окружающей среды. При прочих равных условиях они определяются *только температурой тела.*

---

[1] Здесь мы отступаем от традиции. По историческим причинам излучающую способность тела принято характеризовать *полным потоком* лучистой энергии, посылаемым излучающей площадкой ds наружу по *всем направлениям*, т.е. в пределах телесного угла 2π. При более детальном рассмотрении надо разложить ее по разным направлениям и частотам, что мы и делам.

2. Установим связь между излучательной и поглощательной способностью тела. Так как эти величины характеризуют только поверхность самого тела и совершенно не зависят от окружающего излучения, то в рассуждениях относительно этого излучения можно вводить любые предположения. Предположим, что излучающее тело со всех сторон окружено равновесным излучением, температура которого равна температуре тела. Это, например, можно осуществить, если в качестве рассматриваемого участка поверхности тела взять часть внутренней поверхности замкнутой полости, температура стенок которой поддерживается постоянной. Выделим из всего излучения часть, заполняющую интервал частот ω, $\omega + d\omega$, и рассмотрим превращения ее при излучении и отражении от стенок полости. На площадку ds стенки (рис. 338) за время dt в пределах телесного угла $d\Omega$ падает лучистый поток

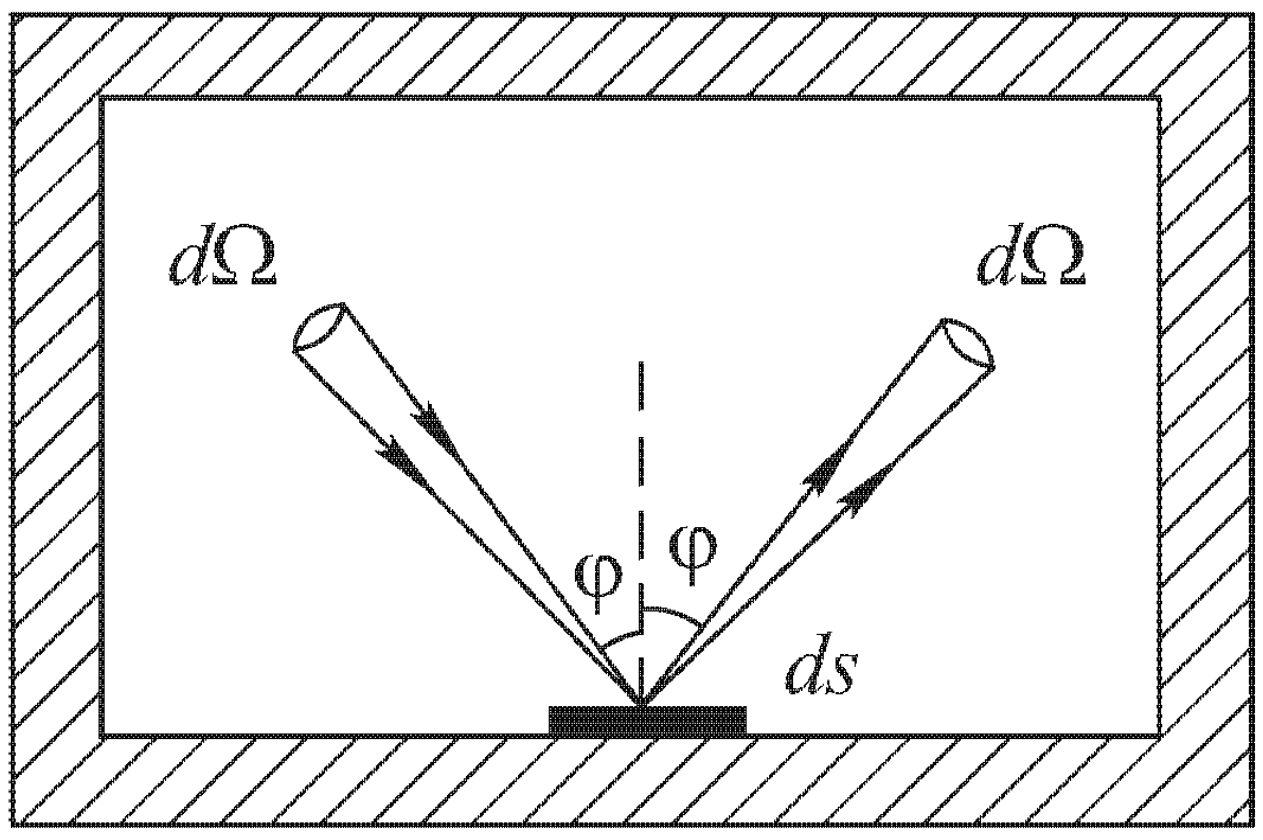


Рис. 338

$$I_\omega\, ds \cos\varphi\, d\omega\, d\Omega\, dt.$$

Часть его

$$(1 - A_\omega) I_\omega\, ds \cos\varphi\, d\omega\, d\Omega\, dt$$

отражается, остальная часть поглощается. На отраженный поток накладывается поток

$$E_\omega\, ds \cos\varphi\, d\omega\, d\Omega\, dt$$

собственного излучения площадки. Таким образом, от площадки ds внутрь полости исходит лучистый поток

$$[(1 - A_\omega) I_\omega + E_\omega]\, ds \cos\varphi\, d\omega\, d\Omega\, dt.$$

Абсолютно черных тел, как и других идеализированных объектов, в природе не существует. Наилучшим приближением к абсолютно черному телу является замкнутая полость, в стенке которой сделано малое отверстие, через которое излучение из полости может выходить наружу. Если стенки полости непрозрачны, то при достаточно малых размерах отверстия в полости установится излучение, лишь бесконечно мало отличающееся от равновесного. Через отверстие будет выходить практически такое же излучение, какое испускалось бы абсолютно черной площадкой той же формы и размеров.

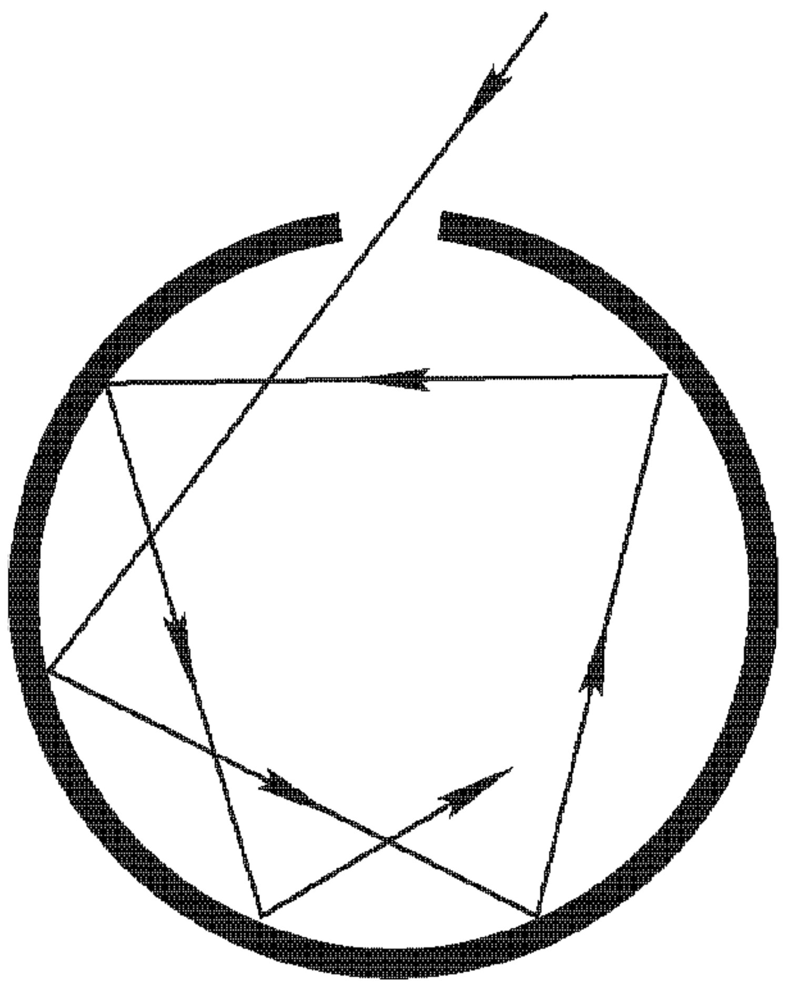


Рис. 339

Убедиться в этом можно и другим способом. Луч света, вступивший снаружи в полость через ее отверстие, будет претерпевать многократные отражения от стенок полости (рис. 339) При каждом отражении часть лучистой энергии поглощается. После многократных отражений луч либо совсем не выйдет наружу через отверстие, либо выйдет лишь ничтожная часть лучистой энергии, вступившей в полость. Почти вся энергия поглотится стенками полости. Это значит, что полость с малым отверстием в отношении поглощения, а потому по закону Кирхгофа и в отношении испускания, ведет себя практически *как абсолютно черное тело.* Изложенный способ всегда применяется при точных количественных измерениях излучения абсолютно черного тела.

Поясним изложенное примером. Если стенки полости с малым отверстием ярко осветить снаружи, то отверстие будет выделяться своей чернотой на светлом фоне освещенных стенок. Таким представляется, например, открытое окно здания. Так будет происходить даже в том случае, когда наружная поверхность стенок полости закрашена черной краской. Если же раскалить стенки полости, сделанные из материала с малой поглощательной способностью (например, из белого фарфора), то отверстие будет ярко светиться на более тусклом фоне прямого излучения самих стенок. При закрашивании стенок снаружи в черный цвет стенки светятся ярче, но все же их яркость остается меньше яркости отверстия. Закрашивание стенок изнутри совсем не сказывается на яркости отверстия.

4. Так как для абсолютно черного тела $e_\omega = I_\omega$, а равновесное излучение изотропно, то излучательная способность абсолютно черного тела одинакова по всем направлениям. При тепловом излучении она, очевидно, совпадает с поверхностной яркостью тела (см. § 22). Значит, *излучательная способность абсолютно черного тела подчиняется закону Ламберта* (см. § 22).

Обратное, конечно, не справедливо. Закон Ламберта может выполняться и для не абсолютно черных тел. Закон Кирхгофа позволяет полнее исследовать этот вопрос. Согласно этому закону, $E_\omega = A_\omega e_\omega$, а потому тело будет излучать по закону Ламберта, если его поглощательная способность не зависит от направления поглощаемого излучения. Так как для непрозрачного тела энергия падающего света равна сумме энергий поглощенного и рассеянного света, то, очевидно, закону Ламберта будет подчиняться излучение всяких тел, равномерно рассеивающих падающий свет во все стороны, независимо от его направления. Такие тела называются *абсолютно матовыми*. Итак, *тепловое излучение абсолютно матовых тел следует закону Ламберта.* На эту связь закона Ламберта с законом Кирхгофа указал В.А. Ульянин (1863–1930) в 1897 г., хотя его рассуждения и были несколько сложнее приведенных здесь.

§ 114. Формула Кирхгофа–Клаузиуса

Найдем теперь плотность энергии и удельную интенсивность равновесного излучения в прозрачной однородной изотропной среде с показателем преломления n. Такое излучение устанавливается в замкнутой полости, заполненной рассматриваемой средой, стенки которой поддерживаются при постоянной температуре. В точности такое же излучение установится в среде и в том случае, когда она заполняет только часть полости. Предположим, что часть полости заполнена рассматриваемой средой, а в другой находится вакуум. Равновесное излучение как в среде, так и в вакууме совершенно не зависит от формы и свойств поверхности, вдоль которой среда граничит с вакуумом. Не меняя окончательного результата, можно принять, что эта граница плоская и гладкая.

Излучение в среде не зависит и от ее размеров. Поэтому среду можно считать настолько протяженной, чтобы при сколь угодно малом коэффициенте поглощения световой луч, вступивший в среду, успел полностью поглотиться, не достигнув стенок полости. Тогда обмен энергией между средой и вакуумом будет происходить только в результате отражения и преломления излучения на рассматриваемой границе. Такой обмен подчиняется *принципу детального равновесия* и не может нарушить состояние равновесия излучения как в среде, так и в вакууме. Из этого условия и можно найти соотношение между плотностью энергии u^0 и удельной интенсивностью I^0 излучения в вакууме с такими же величинами u и I в среде. (В этом параграфе все величины, относящиеся к вакууму, снабжены нулем в индексе, а все величины в среде оставлены без индекса.)

Ввиду выполнения принципа детального равновесия, рассуждение достаточно провести не для всего излучения, а только для произвольной части его, заполняющей интервал частот ω, $\omega + d\omega$. Через единичную площадку границы раздела в единицу времени из вакуума в пределах телесного угла Ω_0 падает лучистый поток $\Phi_1 = I_\omega^0 \, d\Omega_0 \, d\omega \cos\varphi_0$, где φ_0 — угол падения (рис. 340). Согласно принципу детального равновесия, должен существовать такой же лучистый поток, распространяющийся в обратном направлении. Он состоит из двух потоков. Первый поток возникает в результате отражения потока Φ_2 и равен $(1 - A_\omega) I_\omega^0 \, d\Omega_0 \, d\omega \cos\varphi_0$. Второй поток возникает в результате преломления потока Φ_3, приходящего снизу из среды. Так как в силу формул Френеля коэффициенты отражения на границе раздела при прямом и обратном ходе лучей одинаковы, то второй поток будет $A_\omega I_\omega \, d\Omega \, d\omega \cos\varphi$, где φ — угол преломления, а $d\Omega$ — телесный угол в среде, переходящий при преломлении в $d\Omega_0$. Сокращая на $d\omega$, запишем условие детального равновесия в виде

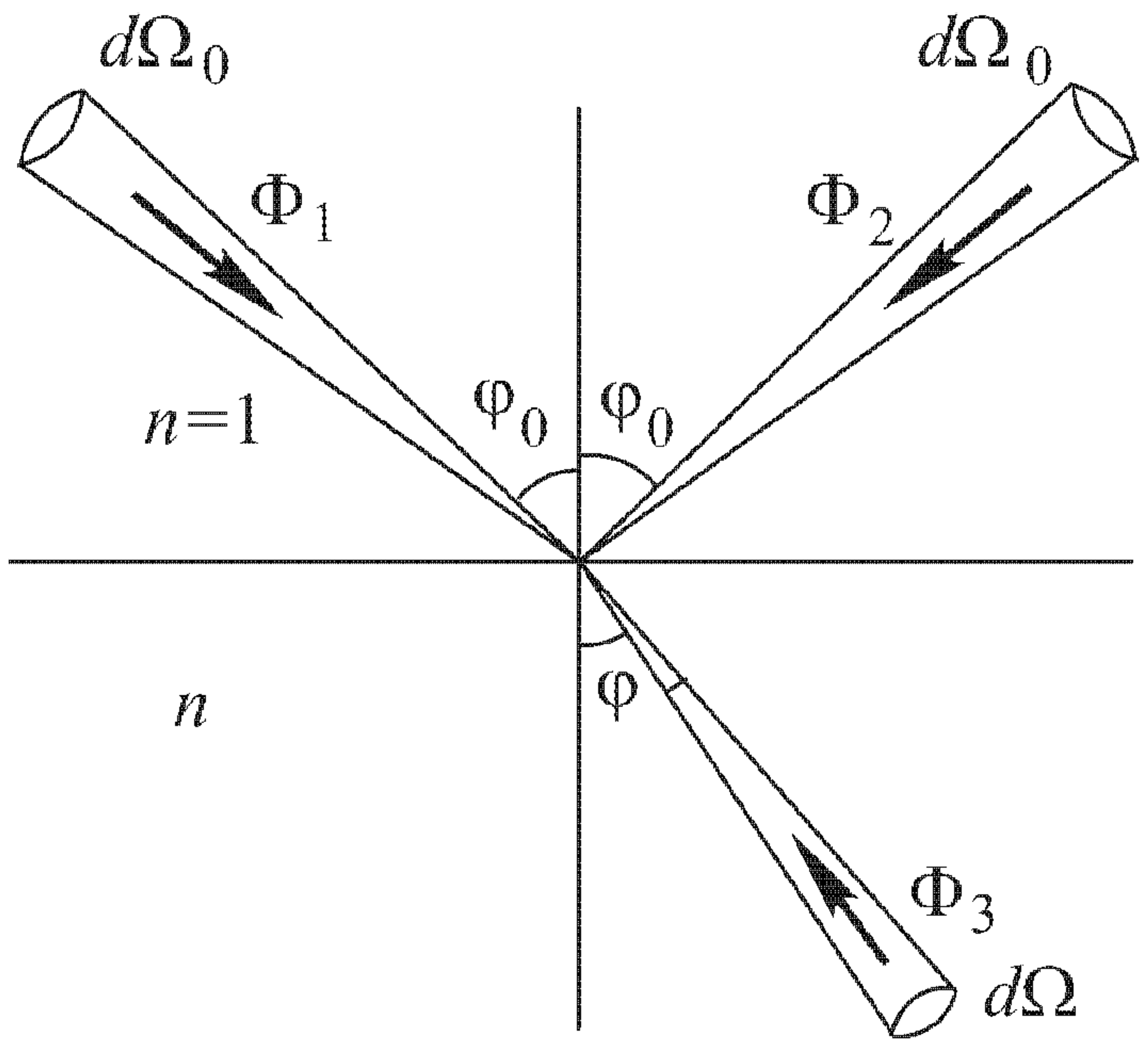


Рис. 340

$$(1 - A_\omega) I_\omega^0 \cos\varphi_0 \, d\Omega_0 + A_\omega I_\omega \cos\varphi \, d\Omega = I_\omega^0 \cos\varphi_0 \, d\Omega_0,$$

или

$$I_\omega \cos\varphi \, d\Omega = I_\omega^0 \cos\varphi_0 \, d\Omega_0.$$

В качестве $d\Omega_0$ возьмем (не изображенный на рис. 340) телесный угол, заключенный между коническими поверхностями, образующие которых составляют с нормалью к границе раздела углы φ_0 и $\varphi_0 + d\varphi_0$, т.е. $d\Omega_0 = 2\pi \sin\varphi_0 \, d\varphi_0$. Аналогично, $d\Omega = 2\pi \sin\varphi \, d\varphi$. Тогда

$$I_\omega \cos\varphi \sin\varphi \, d\varphi = I_\omega^0 \cos\varphi_0 \sin\varphi_0 \, d\varphi_0.$$

В силу закона преломления $\sin\varphi_0 = n \sin\varphi$, так что $\cos\varphi_0 \, d\varphi_0 = n \cos\varphi \, d\varphi$. Поэтому

$$I_\omega \sin\varphi = n I_\omega^0 \sin\varphi_0.$$

Отсюда в силу того же закона преломления

$$I_\omega = n^2 I_\omega^0. \tag{114.1}$$

Эта формула и решает поставленную задачу. Она была получена Кирхгофом в 1860 г. и независимо от него Клаузиусом в 1864 г. Ее называют *формулой Кирхгофа–Клаузиуса*.

В вакууме величины I_ω^0 и u_ω^0 связаны соотношением (112.6), т.е. $cu_\omega^0 = 4\pi I_\omega^0$. В среде оно заменяется на $v_{\text{гр}}u_\omega = 4\pi I_\omega$, где $v_{\text{гр}}$ — групповая скорость. Следовательно,

$$u_\omega = u_\omega^0 n^2 \frac{c}{v_{\text{гр}}}. \tag{114.2}$$

Для недиспергирующих сред

$$u_\omega = n^3 u_\omega^0. \tag{114.3}$$

Соотношениям (114.1) и (114.3) можно придать другую, легче запоминаемую форму, глубокий смысл которой выяснится при рассмотрении проблемы методами статистической физики (см § 117). Для этого заметим, что $n = \lambda_0/\lambda$. Тогда (114.1) и (114.3) преобразуются в

$$\lambda^2 I_\omega = \lambda_0^2 I_\omega^0, \tag{114.4}$$

$$\lambda^3 u_\omega = \lambda_0^3 u_\omega^0. \tag{114.5}$$

Соотношение (114.4) означает, что *в равновесном излучении потоки лучистой энергии через квадратную площадку со стороной, равной длине волны* λ *(в рассматриваемой среде), отнесенные к единице телесного угла, в направлении нормали к площадке одинаковы для всех сред и зависят только от их температуры.* Это утверждение справедливо как для полного потока, так и для его монохроматических составляющих. Аналогично, соотношение (114.5) утверждает, что *энергия равновесного излучения, локализованная в кубике с ребром, равным длине волны* λ, *одинакова во всех изотропных недиспергирующих средах и определяется только температурой среды.* Это утверждение относится также не только к полной плотности энергии, но и к ее монохроматическим составляющим.

§ 115. Закон Стефана–Больцмана

1. В XIX веке производились многочисленные исследования зависимости *интегральной лучеиспускательной способности* нагретых тел от температуры, т.е. величины, которая определяет *суммарную энергию* всех длин волн, излучаемых телами. Эти исследования приводили к противоречивым результатам. Основная причина расхождений была окончательно выяснена после установления закона Кирхгофа, так как излучение определяется не только температурой, но также составом тела и физическими свойствами излучающей поверхности. А на эту сторону дела в экспериментальных исследованиях не обращалось должного внимания. Из эмпирически установленных законов следует отметить только результат, найденный в 1879 г.

Стефаном (1835–1898). Он нашел, что *для черных тел излучательная способность пропорциональна четвертой степени температуры.* Через пять лет Больцман получил этот результат теоретически из термодинамических соображений и показал, что он *абсолютно верен для абсолютно черных тел.* Этот результат, получивший название *закона Стефана–Больцмана*, был подтвержден последующими опытами по излучению абсолютно черного тела.

Вывод Больцмана и все последующие работы по теории теплового излучения существенно используют результаты Максвелла, предсказавшего и рассчитавшего давление света (см. т. II, § 61; т. III, § 145, а также задачу 2 к § 84 этого тома). Для изотропного излучения это давление равно $\mathcal{P} = {}^1\!/_3 u$, где u — интегральная плотность лучистой энергии. К такому выражению должна приводить всякая релятивистская теория света, независимо от того, является ли она корпускулярной или волновой. До теории относительности этот результат, разумеется, не был известен, а результаты Максвелла не считались общепризнанными. В частности, согласно нерелятивистской корпускулярной теории должно было бы быть $\mathcal{P} = {}^2\!/_3 u$, как это предсказывает кинетическая теория газов (см. т. II, § 59). Поэтому опыты П.Н. Лебедева, впервые измерившего в 1900 г. световое давление, подтвердившие результаты Максвелла, имели основополагающее значение для всей термодинамики лучистой энергии.

2. Доказательство закона Стефана–Больцмана мы проведем *методом циклов*, так как таким путем попутно будут получены важные соотношения, используемые в дальнейшем. Допустим, что изотропное излучение произвольного спектрального состава заключено в адиабатическую оболочку с абсолютно зеркальными стенками. Произведем над ним адиабатический квазистатический процесс, при котором объем V, ограниченный оболочкой, меняется бесконечно медленно. Чтобы быть уверенным, что во время этого процесса излучение все время остается изотропным, можно взять оболочку сферической формы (в следующем параграфе будет показано, что эта предосторожность является лишней). Внутренняя энергия излучения в оболочке равна uV. При увеличении объема оболочки на dV за счет этой энергии совершается работа $\mathcal{P}\,dV$, так что $\mathcal{P}\,dV = -d(uV)$. А так как для изотропного излучения $\mathcal{P} = {}^1\!/_3 u$, то этому уравнению можно придать вид ${}^4\!/_3 u\,dV + V\,du = 0$. Отсюда следует, что во время процесса

$$uV^{4/3} = \mathrm{const}, \tag{115.1}$$

или

$$\mathcal{P}V^{4/3} = \mathrm{const}. \tag{115.2}$$

Это — *уравнение адиабаты* для изотропного излучения, совершенно аналогичное уравнению адиабаты Пуассона для идеального газа. Постоянная адиабаты равна $\gamma = {}^4\!/_3$.

В силу эффекта Доплера при адиабатическом сжатии или расширении излучения должен меняться его спектральный состав. Допустим, например, что изотропное излучение занимает спектральный интервал ω, $\omega + d\omega$. В результате отражения от движущейся стенки частота ω и ширина интервала $d\omega$ изменятся и сделаются равными ω' и $d\omega'$. При этом будет выполняться соотношение

$$u_\omega \, d\omega \cdot V^{4/3} = u'_{\omega'} \, d\omega' \cdot V'^{4/3} = \text{const}, \tag{115.3}$$

где V' и $u'_{\omega'}$ — объем и спектральная плотность энергии излучения частоты ω' в конце процесса.

3. Докажем теперь закон Стефана–Больцмана. Для этого произведем над черным излучением *цикл Карно*. Допустим, что излучение заключено в цилиндре, боковые стенки и поршень которого идеально отражающие, а дно черное и может приводиться в тепловой контакт с нагревателем температуры T_1 и холодильником температуры T_2. Излучение можно адиабатически изолировать с помощью идеально отражающей задвижки, вводимой сбоку для прикрытия черного дна цилиндра. В отсутствие задвижки, когда черное дно цилиндра приведено в тепловой контакт с нагревателем или холодильником, излучение в цилиндре, конечно, будет равновесным. Чтобы быть уверенным, что оно сохранится равновесным и во время адиабатического процесса, когда задвижка вставлена, введем внутрь цилиндра черную пылинку, роль которой была выяснена в п. 1 § 112. На изотерме *1–2* (рис. 341) дно цилиндра контактирует с нагревателем. Количество тепла, переданное нагревателем на этой изотерме, равно

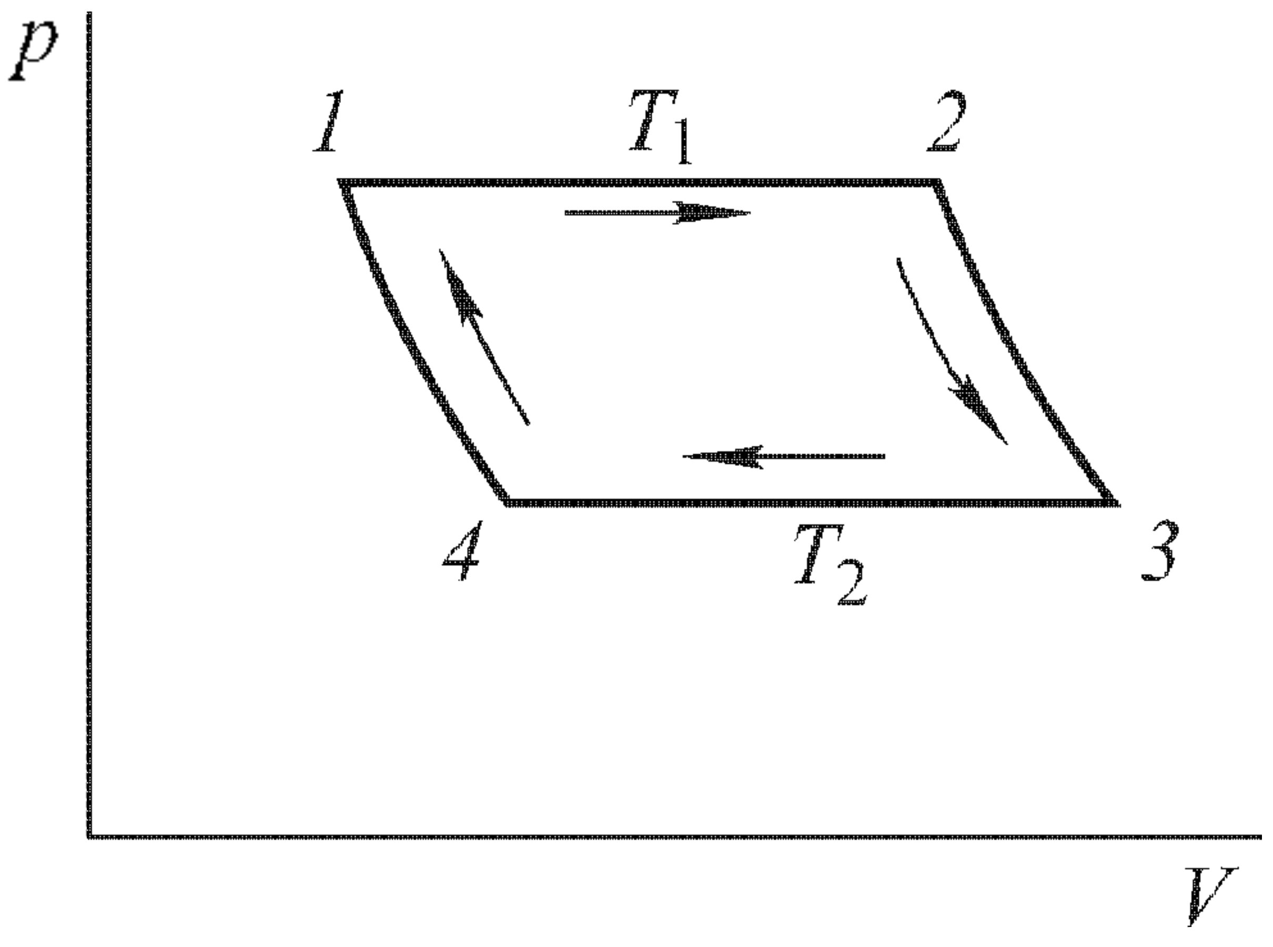


Рис. 341

$$Q_1 = u_1(V_2 - V_1) + \mathcal{P}_1(V_2 - V_1) = \frac{4}{3} u_1(V_2 - V_1).$$

Количество тепла, отданное холодильнику на изотерме *3–4*, $Q_2 = {}^4\!/_3 u_2(V_3 - V_4)$. По теореме Карно

$$\frac{Q_1}{Q_2} = \frac{u_1(V_2 - V_1)}{u_2(V_3 - V_4)} = \frac{T_1}{T_2}.$$

На адиабатах *2–3* и *4–1* в силу (115.1) выполняются соотношения

$$u_1^{3/4} V_2 = u_2^{3/4} V_3, \quad u_1^{3/4} V_1 = u_2^{3/4} V_4.$$

Отсюда

$$\frac{V_2 - V_1}{V_3 - V_4} = \left(\frac{u_2}{u_1}\right)^{3/4},$$

а потому $u_1/T_1^4 = u_2T_2^4 = \mathrm{const}$. Следовательно,

$$u = aT^4, \tag{115.4}$$

где a — универсальная постоянная. Но это есть иная форма закона Стефана–Больцмана.

Результат (115.4) можно получить короче, если к равновесному излучению применить общую термодинамическую формулу

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial \mathcal{P}}{\partial T}\right)_V - \mathcal{P}. \tag{115.5}$$

Подставив сюда $U = Vu(T)$, $\mathcal{P} = {}^1\!/_3u(T)$, придем к тому же дифференциальному уравнению, интегрированием которого была получена формула (115.4). Однако мы не хотели пользоваться формулой (115.5).

§ 116. Теорема и закон смещения Вина

1. Важные результаты в термодинамике излучения были получены Вильгельмом Вином (1864–1928) в 1893–1894 гг. Вин доказал, что *равновесное излучение, заключенное в оболочке с идеально отражающими стенками, будет оставаться равновесным при квазистатическом сжатии или расширении оболочки.*

Для наших целей при доказательстве теоремы Вина достаточно ограничиться оболочкой сферической формы. В этом случае, ввиду сферической симметрии системы, отпадает необходимость специально доказывать, что в ходе процесса изотропия излучения все время сохраняется. Сожмем излучение квазистатически от начального объема V_1 до конечного V_2. При этом будет совершена работа против сил светового давления, и энергия излучения в оболочке увеличится. Спектральный состав излучения также изменится, из-за эффекта Доплера. Допустим, что в результате этого излучение перестанет быть равновесным. Введем внутрь оболочки в конечном состоянии бесконечно малую черную пылинку, поглощающую и излучающую свет. По истечении достаточно длительного времени она превратит неравновесное излучение в оболочке в равновесное. Это — необратимый процесс, идущий самопроизвольно. Обратный процесс превращения равновесного излучения в неравновесное, разумеется, сам собою идти не может.

Когда излучение внутри оболочки станет равновесным, не убирая пылинку, начнем бесконечно медленно адиабатически расширять оболочку, доведя объем излучения до исходного значения V_1. После этого удалим пылинку. Энергия пылинки бесконечно мала, а потому ее наличие может сказаться на общей энергии излучения в полости также бесконечно мало. С другой стороны, давление изотропного излучения зависит только от интегральной плотности энергии излучения u, но не от его спектрального состава Поэтому работа, которую совершит световое давление при расширении оболочки, будет с точностью до бесконечно малой величины равна внешней работе, совершенной над излучением при его сжатии. Отсюда следует, что в результате сжатия и последующего расширения энергия, а с ней и температура излучения не изменятся.

Система совершила круговой процесс, в ходе которого она не получала и не отдавала тепло, а общая работа, произведенная ею, равна нулю. Значит, в окружающих телах не произошло никаких изменений, а потому рассматриваемый круговой процесс обратим. Но это невозможно, так как одна из стадий этого кругового процесса по нашему предположению необратима. Следовательно, это предположение неверно, и теорема Вина доказана.

2. Значение теоремы Вина — методическое. Действительно, адиабатически и квазистатически меняя объем равновесного излучения в оболочке с идеально зеркальными стенками, можно получить равновесное излучение произвольной плотности, а следовательно, и температуры. Энергию (и температуру) этого излучения можно найти, вычислив работу, совершенную над ним в этом процессе. Его спектральный состав найдется, если вычислить доплеровское изменение частоты излучения при его отражении от движущейся оболочки. Тем самым будет установлено определенное соответствие между параметрами равновесного излучения в начале процесса и на любой стадии его.

Применим этот метод к равновесному излучению в сферической оболочке с идеально зеркальными стенками. При бесконечно медленном адиабатическом расширении или сжатии оболочки излучение в ней все время будет оставаться равновесным, так что его можно в любой момент времени характеризовать определенной температурой T. Выделим внутри оболочки произвольный луч, падающий на оболочку под углом ϑ (рис. 342). Время между двумя последовательными отражениями этого луча равно $\Delta t = (2r/c)\cos\vartheta$. За это время радиус оболочки r полу-

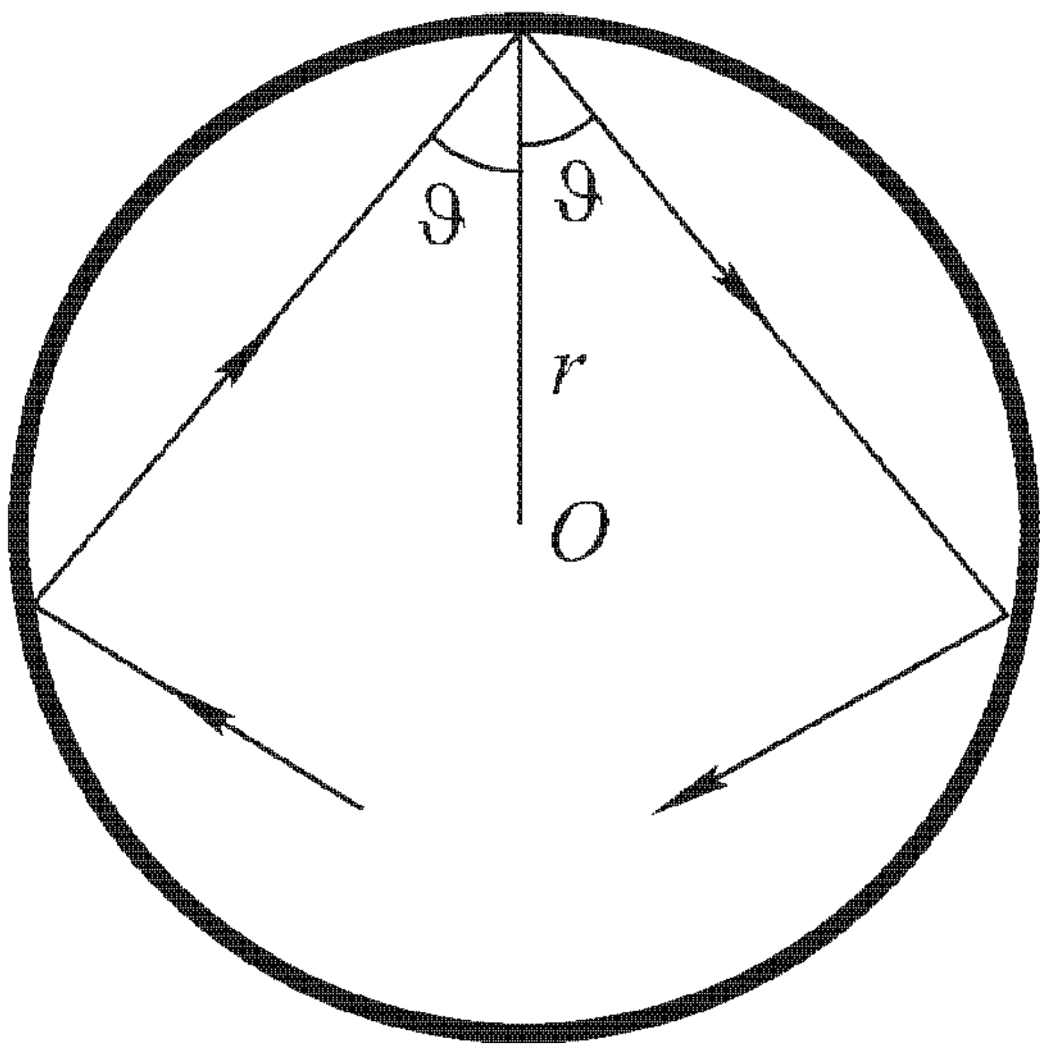


Рис. 342

чает приращение $\Delta r = \dot{r}\,\Delta t$. При каждом отражении происходит доплеровское изменение частоты, определяемое формулой

$$\frac{\Delta\omega}{\omega} = -\frac{2\dot{r}\cos\vartheta}{c} = -\frac{2\Delta r\cos\vartheta}{c\,\Delta t} = -\frac{\Delta r}{r},$$

если пренебречь квадратом бесконечно малой радиальной скорости $\dot{r}$ расширения оболочки. Относительное изменение частоты $\Delta\omega/\omega$ определяется только относительным изменением $\Delta r/r$ радиуса оболочки. Такая же формула получится и в том случае, когда за время изменения радиуса оболочки на Δr произойдет не одно, а много отражений светового луча. Требуется только, чтобы выполнялось условие $\Delta r \ll r$. При бесконечно медленном расширении величины Δr и $\Delta\omega$ можно заменить их дифференциалами, т.е. написать

$$\frac{d\omega}{\omega} + \frac{dr}{r} = 0. \qquad (116.1)$$

Это означает, что реальный процесс, в котором последовательные отражения отделены друг от друга малыми, но все же конечными промежутками времени, при расчетах заменяется идеализированной схемой, в которой эти отражения следуют друг за другом *непрерывно во времени*. Интегрируя уравнение (116.1), получим

$$\omega r = \mathrm{const}. \qquad (116.2)$$

Так как $r \sim V^{1/3}$, то этот результат можно записать также в виде

$$\omega^3 V = \mathrm{const}. \qquad (116.3)$$

В таком виде он справедлив для полости произвольной формы. А поскольку он получен для бесконечно медленного процесса, величина $\omega^3 V$ является *адиабатическим инвариантом*. Комбинируя его с ранее полученными адиабатическими инвариантами (115.1) и (115.3), получим новые адиабатические инварианты. Так, из формул (115.1) и (116.3) следует

$$\frac{\omega^4}{u} = \mathrm{const}, \qquad (116.4)$$

или на основании закона Стефана–Больцмана

$$\frac{\omega}{T} = \mathrm{const}. \qquad (116.5)$$

Аналогично, формулы (115.3) и (116.3) дают

$$\frac{u_\omega\,d\omega}{\omega^4} = \mathrm{const}. \qquad (116.6)$$

Таким образом, при квазистатическом расширении или сжатии равновесного излучения в полости с зеркальными стенками каждая квазимонохроматическая составляющая излучения ведет себя *независимо от остальных составляющих* и меняется так, что величины $\omega^3 V$, u/ω^4 и $u_\omega\, d\omega/\omega^4$ остаются постоянными, т.е. являются адиабатическими инвариантами. По теореме Вина при таком процессе излучение все время остается равновесным. Такое же излучение можно было бы получить в неподвижной оболочке, нагревая или охлаждая ее стенки. Поэтому полученные результаты можно представить как *свойства только самого равновесного излучения*, не связывая их ни с каким конкретным процессом. Сформулируем их следующим образом. *Изменим любым способом температуру равновесного излучения от T до T', чтобы излучение оставалось равновесным. Каждой частоте ω излучения в начальном состоянии приведем в соответствие такую частоту ω' в конечном состоянии, чтобы $\omega/T = \omega'/T'$ и, следовательно, $d\omega/T = d\omega'/T'$. Тогда плотности лучистой энергии в этих состояниях будут связаны соотношениями*

$$\frac{u}{\omega^4} = \frac{u'}{\omega'^4}, \tag{116.7}$$

$$\frac{u_\omega\, d\omega}{\omega^4} = \frac{u'_{\omega'}\, d\omega'}{\omega'^4}. \tag{116.8}$$

Эти результаты составляют содержание так называемого *закона смещения Вина в его наиболее общей форме.*

3. Из формулы (116.8) получаем

$$u_\omega(\omega, T) = \frac{\omega^4}{\omega'^4}\frac{d\omega'}{d\omega}u'_{\omega'}(\omega', T) = \frac{T^3}{T'^3}u'_{\omega'}\left(\frac{T'}{T}\,\omega, T'\right).$$

Это соотношение справедливо при любом значении температуры T', а потому величина справа от T' не зависит. Величине T' можно придать *любое значение*, представив полученное соотношение в виде

$$u_\omega(\omega, T) = T^3\varphi\left(\frac{\omega}{T}\right), \tag{116.9}$$

где $\varphi(\omega/T)$ — *универсальная функция* аргумента ω/T. Ввиду соотношения $\omega/T = \omega'/T'$, тот же результат можно записать в виде

$$u_\omega(\omega, T) = \omega^3 f\left(\frac{\omega}{T}\right), \tag{116.10}$$

где $f(\omega/T) \equiv (T'/\omega')^3\varphi(\omega/T)$ — новая *универсальная функция* того же аргумента ω/T. Тем самым определение универсальной функции $u_\omega(\omega, T)$ *двух аргументов* сведено к задаче нахождения универсальной функции $\varphi(\omega, T)$ или $f(\omega/T)$ только *одного аргумента* ω/T. Отсюда следует, что *если известно спектральное распределение в равновесном излучении при какой-либо произвольной температуре T', то с помощью формулы* (116.9) *или* (116.10) *можно найти это распределение при всякой другой температуре T.*

В переменных λ, T, как это следует из соотношения (112.2), формулам (116.9) и (116.10) можно придать вид

$$u_\lambda = T^5 \varphi_1(\lambda T), \tag{116.11}$$

$$u_\lambda = \frac{1}{\lambda^5} f_1(\lambda T), \tag{116.12}$$

где $\varphi_1(\lambda T)$ и $f_1(\lambda T)$ — новые *универсальные функции*. При фиксированной температуре T величина u_λ становится функцией только длины волны λ. Эта функция не может возрастать монотонно, а должна иметь максимум. В противном случае интегральная плотность излучения

$$u = \int\limits_0^\infty (u_\lambda)_{T=\mathrm{const}}\, d\lambda \tag{116.13}$$

не могла бы оставаться конечной. Длину волны в максимуме обозначим через λ_m. Если ввести обозначение $x = \lambda T$, то для определения положения максимума получится уравнение

$$d\varphi_1/d\lambda = T\, d\varphi_1/dx = 0, \quad \text{т.е.} \quad d\varphi_1/dx = 0.$$

Таким образом, при всех температурах максимум получается при одном и том же значении аргумента x. Отсюда следует, что *при повышении температуры максимум функции* $u_{\lambda, T=\mathrm{const}}$ *смещается в сторону более коротких волн и притом так, что выполняется соотношение*

$$\lambda_m T = b = \mathrm{const}. \tag{116.14}$$

Измерения дают: $b = 0{,}2898$ см·К. Этот результат, определяющий смещение максимума излучения при изменении температуры T, называют *законом Вина в его специальной форме.* В точках максимума функции $u_\lambda(T)$ и $u_{\lambda'}(T')$ относятся как *пятые степени* абсолютных температур T и T'.

Функция $u_\omega(\omega, T)$ при постоянной температуре T также обращается в максимум при какой-то частоте $\omega = \omega_m$. Частота ω_m не равна $2\pi c/\lambda_m$, так как речь идет о максимумах *различных функций* $u_\omega(\omega)$ и $u_\lambda(\lambda)$. При изменении температуры излучения положение максимума смещается, но при этом имеет место соотношение

$$\frac{\omega_m}{T} = \mathrm{const}, \tag{116.15}$$

которое является *другой формой закона смещения Вина* (116.14).

4. Выведенный ранее закон Стефана–Больцмана является следствием общей формулы (116.9). Действительно, вводя обозначение $x = \omega/T$, из этой формулы при постоянной T находим

$$u = \int\limits_0^\infty u_\omega\, d\omega = T^3 \int\limits_0^\infty \varphi\left(\frac{\omega}{T}\right)\, d\omega = T^4 \int\limits_0^\infty \varphi(x)\, dx = aT^4,$$

где $a \equiv \int\limits_0^\infty \varphi(x)\, dx$ — универсальная постоянная.

§ 117. Формула Рэлея–Джинса

1. Результатами, изложенными в предыдущих параграфах, исчерпывается все, что могла дать феноменологическая термодинамика в проблеме теплового излучения. Ее оказалось недостаточно для решения основной проблемы теории теплового излучения: определения функции $u_\omega(\omega, T)$ или функции $I_\omega(\omega, T)$, связанной с ней соотношением (112.6). Для этого оказалось необходимым привлечь *статистические методы* и учесть *квантовые свойства вещества и излучения*. Первая попытка теоретического решения указанной проблемы была предпринята в 1887 г. В.А. Михельсоном (1860–1927). В то время, как показало последующее развитие физики, правильное решение рассматриваемой проблемы было, конечно, невозможно. Заслуга Михельсона состоит в том, что он привлек внимание физиков к одной из важнейших проблем, решение который положило начало *квантовой физики*.

Общий метод теоретического определения функции $u_\omega(\omega, T)$ в рамках классической физики, не связанный с модельными представлениями, был указан в 1900 г. Рэлеем и через пять лет более подробно развит Джинсом (1877–1946). Рэлей и Джинс применили к равновесному излучению в полости теорему классической статистической механики о *равномерном распределении кинетической энергии по степеням свободы*. Согласно этой теореме, *в состоянии статистического равновесия на каждую степень свободы приходится в среднем кинетическая энергия* $^1\!/_2 kT$, где $k = 1{,}38 \cdot 10^{-16}$ эрг/К — постоянная Больцмана.

Если степень свободы колебательная, то надо учесть еще *потенциальную энергию*. В случае *гармонических колебаний* среднее значение потенциальной энергии равно также $^1\!/_2 kT$ (см. т. II, § 63). Таким образом, *в состоянии статистического равновесия на каждую колебательную степень свободы приходится средняя энергия, равная* kT.

Эта теорема сводит задачу нахождения функции $u_\omega(\omega, T)$ к определению *числа степеней свободы излучения в полости*. Поскольку равновесное излучение в полости не зависит от ее формы и материала стенок, можно предположить, что полость имеет форму куба с идеально отражающими стенками. А чтобы излучение в полости было равновесным, можно ввести в нее бесконечно малую черную пылинку, как это делалось в предыдущем параграфе при доказательстве теоремы Вина.

2. Чтобы лучше уяснить метод определения числа степеней свободы, рассмотрим этот вопрос сначала не для векторного электромагнитного поля, а для *скалярного волнового поля*, например для продольных акустических волн. В такой постановке этот вопрос имеет и самостоятельный интерес, например в теории теплоемкости твердых тел Дебая. Волновое поле будем характеризовать какой-то функцией $V(\mathbf{r}, t)$, удовлетворяющей волновому уравнению

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = 0. \tag{117.1}$$

Предположим, что на стенках полости функция V обращается в нуль. Тогда (а также и при других граничных условиях) по теореме Фурье функция $V(\mathbf{r}, t)$ может быть представлена суперпозицией стоячих волн. Координатные оси X, Y, Z направим параллельно ребрам кубической полости. Так как граничные условия должны удовлетворяться на гранях $x = 0$, $y = 0$, $z = 0$ и на параллельных им гранях полости, то каждая стоячая волна должна представляться функцией с *разделяющимися переменными*, т.е.

$$V = X(x)Y(y)Z(z)\exp(i\omega t).$$

Отсюда

$$\frac{1}{V}\frac{\partial^2 V}{\partial x^2} = \frac{1}{X}\frac{d^2 X}{dx^2}, \ldots, \frac{1}{V}\frac{\partial^2 V}{\partial t^2} = \frac{\omega^2}{c^2}.$$

Подставив эти выражения в уравнение (117.1), получим

$$\frac{1}{X}\frac{d^2 X}{dx^2} + \frac{1}{Y}\frac{d^2 Y}{dy^2} + \frac{1}{Z}\frac{d^2 Z}{dz^2} = \frac{\omega^2}{c^2}.$$

Левая часть этого уравнения есть сумма функций от x, y, z соответственно, т.е. от различных переменных. Она равна постоянной ω^2/c^2. Это может быть тогда и только тогда, когда каждая из этих функций сама постоянна, т.е.

$$\frac{1}{X}\frac{d^2 X}{dx^2} = q_x^2, \quad \frac{1}{Y}\frac{d^2 Y}{dy^2} = q_y^2, \quad \frac{1}{Z}\frac{d^2 Z}{dz^2} = q_z^2,$$

где q_x, q_y, q_z — постоянные, удовлетворяющие условию

$$q_x^2 + q_y^2 + q_z^2 = q^2 \equiv \frac{\omega^2}{c^2}. \tag{117.2}$$

Решение первого уравнения запишем в виде $X = \sin(q_x x + \delta_x)$, где δx — постоянная интегрирования. Второй постоянной интегрирования служит амплитуда, которую без потери общности можно опустить. На грани полости $x = 0$ функция X должна обращаться в нуль, т.е. $\sin\delta_x = 0$. Также не теряя общности, можно положить $\delta_x = 0$. На противоположной грани $x = l$, где l — длина ребра кубической полости, функция X также должна обращаться в нуль, а потому $\sin k_x l = 0$. Такие же результаты получаются и для координат y и z. Следовательно,

$$V = \sin q_x x \, \sin q_y y \, \sin q_z z \, e^{i\omega t}, \tag{117.3}$$

где

$$q_x l = m_x \pi, \quad q_y l = m_y \pi, \quad q_z l = m_z \pi, \tag{117.4}$$

а m_x, m_y, m_z — целые числа. Все их можно считать положительными, так как введение отрицательных чисел не приводит к новым, линейно независимым решениям. В выражение (117.3) можно было бы ввести еще постоянную амплитуду, зависящую от m_x, m_y, m_z, но для наших целей в этом нет надобности. Это выражение и представляет общий вид *стоячей волны* в полости. Каждой тройке целых положительных чисел m_x, m_y, m_z, удовлетворяющей условиям (117.4), соответствует одна стоячая волна. Число возможных стоячих волн бесконечно велико.

Будем рассматривать q_x, q_y, q_z как прямоугольные координаты точки трехмерного «пространства волновых векторов» $\mathbf{q}$. Эти «изображающие точки» расположатся в узлах кубической решетки, элементарная ячейка которой есть кубик с длиной стороны $\Delta q_x = (\pi/l)\Delta m_x = \pi/l$ и объемом $(\pi/l)^3$. Решетка заполняет только положительный октант пространства волновых векторов, так как все координаты q_x, q_y, q_z положительны. Объем части шара радиуса q, лежащей в этом октанте, равен $^1/_8(4\pi/3)q^3 = (\pi/6)q^3$. Число изображающих точек в нем и будет равно числу Z стоячих волн, волновые числа которых не превосходят $q = 2\pi/\lambda$. Подавляющее число волн очень короткие, для них величина q очень велика по сравнению с длиной π/l ребра элементарного кубика. Поэтому число стоячих волн в указанном октанте шара найдется делением его объема на объем элементарного кубика. Таким путем получается асимптотическая формула

$$Z = \frac{(\pi/6)q^3}{\pi^3/l^3} = \frac{V}{6\pi^2}\, q^3 = \frac{V}{6\pi^2}\frac{\omega^3}{c^3}, \tag{117.5}$$

где $V = l^3$ — объем полости. Она справедлива, когда сторона кубической полости очень велика по сравнению с длиной волны $\lambda = 2\pi/q$. Можно доказать, что формула (117.5) остается верной и для полости произвольной формы, хотя в этом случае стоячие волны и не будут представляться выражениями вида (117.3).

Дифференцируя (117.5), получим

$$dZ = \frac{V}{2\pi^2}\, q^2\, dq = \frac{V}{2\pi^2 c^3}\, \omega^2\, d\omega. \tag{117.6}$$

Эта асимптотическая формула дает число стоячих волн в интервале частот ω, $\omega + d\omega$. Она, разумеется, справедлива только для достаточно широких интервалов $d\omega$, когда $dZ \gg 1$.

3. Рассуждения существенно не изменятся и в случае векторного (электромагнитного) поля. В этом случае вектор $\mathbf{E}$ и его прямоугольные составляющие E_x, E_y, E_z удовлетворяют прежнему волновому уравнению (117.1). Для полости с идеально зеркальными стенками граничные условия требуют обращения в нуль тангенциальных составляющих вектора $\mathbf{E}$ на стенках полости. Для этого необходимо и достаточно, чтобы выполнялись следующие условия:

$$\begin{aligned} E_y = E_z &= 0 \quad \text{при} \quad x = 0 \quad \text{и} \quad x = l, \\ E_z = E_x &= 0 \quad \text{при} \quad y = 0 \quad \text{и} \quad y = l, \\ E_x = E_y &= 0 \quad \text{при} \quad z = 0 \quad \text{и} \quad z = l. \end{aligned}$$

Стоячая волна, удовлетворяющая этим условиям, имеет вид

$$\begin{aligned} E_x &= \sin(q_x x + \varphi_x) \sin q_y y \, \sin q_z z \, e^{i\omega t}, \\ E_y &= \sin q_x x \, \sin(q_y y + \varphi_y) \sin q_z z \, e^{i\omega t}, \\ E_z &= \sin q_x x \, \sin q_y y \, \sin(q_z z + \varphi_z) \, e^{i\omega t}, \end{aligned}$$

где q_x, q_y, q_z определяются прежними формулами (117.4). Что касается фаз φ_x, φ_y, φ_z, то их можно было бы найти из граничных условий, которым на стенках полости удовлетворяют нормальные составляющие электрического вектора $\mathbf{E}$. Но для наших целей в этом нет необходимости, так как для определения величин Z и dZ достаточно знать только значения, которые могут принимать составляющие q_x, q_y, q_z. Однако по сравнению с предыдущим случаем выражения (117.5) и (117.6) надо удвоить, так как электромагнитные волны *векторные* и притом *поперечные*. Каждому направлению распространения соответствуют *две волны*, поляризованные во взаимно перпендикулярных плоскостях, в результате суперпозиции которых может быть получена волна любой поляризации, распространяющаяся в том же направлении. Итак, для электромагнитного поля

$$dZ = \frac{V}{\pi^2 c^3} \omega^2 \, d\omega. \tag{117.7}$$

На каждую стоячую волну в состоянии статистического равновесия приходится в среднем энергия $\bar{\mathcal{E}} = kT$: одна половина ее — электрическая, другая — магнитная. Записав энергию равновесного излучения в полости в спектральном интервале $d\omega$ в виде $Vu_\omega \, d\omega$, из формулы (117.7) получим

$$u_\omega = \frac{\omega^2}{\pi^2 c^3} \bar{\mathcal{E}} = \frac{kT}{\pi^2 c^3} \omega^2. \tag{117.8}$$

Этот результат известен под названием *формулы Рэлея–Джинса*, хотя он независимо и практически одновременно с Рэлеем был получен также Планком из столь же общих, но несколько других соображений (см. п. 6).

4. Прежде чем обсуждать формулу Рэлея–Джинса, заметим, что в случае полости, заполненной изотропной средой, число стоячих волн будет определяться прежними формулами (117.5) и (117.6), если только в них величину c заменить скоростью света v в рассматриваемой среде (предполагается, что среда изотропная). Отсюда следует, что числа Z и dZ в одном и том же интервале частоты, а с ними и функция u_ω пропорциональны c^3/v^3, т.е. кубу показателя преломления среды n. Но это есть *закон Кирхгофа–Клаузиуса*, доказанный в § 114. Вывод справедлив при более общих предположениях, чем это сделано в тексте. Нет необходимости ссылаться на классическую теорему о равномерном распределении энергии по степеням свободы. Достаточно, чтобы средняя энергия гармонического осциллятора была функцией только частоты ω, как это имеет место в квантовой теории.

5. Формула Рэлея–Джинса согласуется с общей термодинамической формулой Вина (116.9) или (116.10). Более того, вид этой формулы может быть непосредственно установлен на основе одной только формулы Вина, если ограничиться предельным случаем низких частот. Для этого заметим, что в формуле (116.10) функция $f(\omega/T)$ должна быть возрастающей функцией температуры T и обращаться в нуль при $T = 0$. Ее удобнее рассматривать как функцию аргумента T/ω и разложить в ряд по степеням этого аргумента. Если ограничиться первым членом этого ряда, то получится

$$u_\omega(\omega, T) = CT\omega^2. \tag{117.9}$$

Но по своему виду этот результат совпадает с формулой (117.8). Только коэффициент C остается неопределенным. Зато формула (117.9) выведена в более общих предположениях, чем (117.8). Таким образом, в предельном случае *низких частот* нет сомнений в правильности вида формулы (117.9). И действительно, для достаточно длинных волн (точный критерий приводится в следующем параграфе) формула Рэлея–Джинса прекрасно согласуется с опытом. Она с успехом применяется в длинноволновой инфракрасной области спектра и в радиодиапазоне.

Но классическая статистика требует, чтобы формула Рэлея–Джинса (117.8) была верна *при любых частотах*. Однако это невозможно, так как тогда для интегральной плотности энергии получилось бы бесконечное значение:

$$u = \frac{kT}{\pi^2 c^3} \int\limits_0^\infty \omega^2\, d\omega = \infty.$$

Отсюда следует, что *по теории Рэлея–Джинса тепловое равновесие между веществом и излучением невозможно.* Этот вывод противоречит опыту. П.С. Эренфест назвал его *ультрафиолетовой катастрофой.* Причина ультрафиолетовой катастрофы заключается в том, что в теории Рэлея–Джинса излучение в полости имеет *бесконечное*, а вещество *конечное* число степеней свободы. Поэтому, если бы было справедливо равномерное распределение энергии по степеням свободы, то при тепловом равновесии вся энергия должна была бы сосредоточиться в излучении.

6. Можно было бы возразить, что классическая статистическая механика, следствием которой является теорема о равномерном распределении кинетической энергии по степеням свободы, неприменима к системам с бесконечным числом степеней свободы. Но такое возражение неубедительно. В основе классической статистической механики лежат уравнения классической механики в форме Гамильтона (1805–1865). Хотя они и были установлены для механических систем с конечным числом степеней свободы, но можно показать, что *излучение в полости можно описывать бесконечным, но счетным числом обобщенных координат, также подчиняющихся уравнениям Гамильтона.* Следовательно, и вся система, состоящая из вещества и излучения, будет описываться уравнениями Гамильтона.

Поэтому было бы непонятно, почему теорема о равномерном распределении энергии справедлива для одних и не имеет места для других степеней свободы.

Кроме того, к формуле Рэлея–Джинса независимо пришел также Планк, применивший эту теорему *только к веществу, но не к излучению.* Он провел рассуждение для одномерного гармонического осциллятора, например, квазиупруго-связанного электрона, помещенного в полость с равновесным излучением. Под действием хаотически меняющегося электромагнитного поля излучения осциллятор будет совершать колебания с хаотически меняющимися амплитудами и фазами, излучая и поглощая при этом электромагнитные волны. Энергия осциллятора будет совершать беспорядочные флуктуации вокруг среднего значения $\bar{\mathcal{E}}$. В результате идейно-простых, но несколько длинных вычислений Планк пришел к формуле

$$u_\omega = \frac{\bar{\mathcal{E}}}{\pi^2 c^3}\, \omega^2. \qquad (117.10)$$

Если в этой формуле применить к осциллятору (вещество!) теорему о равномерном распределении кинетической энергии по степеням свободы, то получится формула Рэлея–Джинса.

§ 118. Формула Планка

1. Правильная формула для спектральной плотности энергии равновесного излучения, подтвержденная всеми экспериментальными исследованиями, была найдена Планком сначала полуэмпирическим путем. Спустя короткое время, Планк нашел теоретический вывод этой формулы, изложенный им 14 декабря 1900 г. на заседании Немецкого физического Общества. Этот день считается днем рождения новой — *квантовой* — физики. Идея о квантах, осторожно формулированная Планком, развилась в стройное и глубокое учение, покорившее всю физику. Решение проблемы равновесного излучения Планком было бы невозможно, если бы в своих исследованиях он не находился в тесном контакте с экспериментаторами. Из экспериментальных работ по излучению абсолютно черного тела особенно следует отметить исследования Лэнглея (1834–1906), Рубенса (1865–1922), Пашена (1865–1947), Вина, Луммера (1860–1925), Э. Принсхейма (1859–1917), Курльбаума.

2. Гипотеза Планка состоит в том, что излучение и поглощение света веществом происходит не непрерывно, а конечными порциями, называемыми *квантами света* или *квантами энергии*. Получим формулу Планка тем же методом, который применялся при выводе формулы Рэлея–Джинса. Тогда гипотезу Планка удобно взять в следующей форме: энергия гармонического осциллятора может принимать не произвольные, а только избранные значения, образующие дискретный ряд: 0, $\mathcal{E}_0$, $2\mathcal{E}_0$, $3\mathcal{E}_0$, ..., где $\mathcal{E}_0$ — определенная величина, зависящая только от собственной частоты ω осциллятора. Здесь под осциллятором понимается не только частица, могущая совершать гармонические свободные колебания, но, например, и стоячая волна определенной частоты в полости.

Если осциллятор изолирован, то по истечении достаточно длительного времени он потеряет всю свою энергию на излучение и перейдет на наинизший энергетический уровень с энергией $\mathcal{E} = 0$. Но если осциллятор находится в полости, стенки которой поддерживаются при постоянной температуре, то наряду с излучением будут происходить и акты поглощения, в результате которых возбуждаются и высшие энергетические уровни. Установится вполне определенное состояние детального равновесия, в котором число актов излучения в среднем равно числу обратных актов поглощения. В этом состоянии будут возбуждены все энергетические уровни, но с различными вероятностями. И все, что требуется для нахождения функции $u_\omega(\omega, T)$, — это определить среднюю энергию $\bar{\mathcal{E}}$ осциллятора в этом состоянии статистического равновесия. Такая задача уже была решена нами (см. т. II, § 85). Приведем еще раз это решение в несколько измененной форме. По теореме Больцмана вероятности возбуждения энергетических уровней осциллятора пропорциональны величинам

$$1, e^{-\mathcal{E}_0/kT}, e^{-2\mathcal{E}_0/kT}, \ldots$$

Поэтому

$$\bar{\mathcal{E}} = \frac{\sum\limits_{n=1}^{\infty} n\mathcal{E}_0 e^{-n\mathcal{E}_0/kT}}{\sum\limits_{n=0}^{\infty} e^{-n\mathcal{E}_0/kT}} = \mathcal{E}_0 \frac{\sum\limits_{n=1}^{\infty} ne^{-nx}}{\sum\limits_{n=0}^{\infty} e^{-nx}},$$

где введено обозначение $x = \mathcal{E}_0/kT$. Значение знаменателя определяется формулой

$$\sum_{n=0}^{\infty} e^{-nx} = \frac{1}{1 - e^{-x}}.$$

Числитель находится дифференцированием этой формулы по x:

$$\sum_{n=1}^{\infty} ne^{-nx} = \frac{e^{-x}}{(1 - e^{-x})^2},$$

и следовательно,

$$\bar{\mathcal{E}} = \frac{\mathcal{E}_0}{e^x - 1} = \frac{\mathcal{E}_0}{e^{\mathcal{E}_0/kT} - 1}. \tag{118.1}$$

Подставив это значение в формулу (117.8), найдем

$$u_\omega = \frac{\omega^2}{\pi^2 c^3} \frac{\mathcal{E}_0}{e^{\mathcal{E}_0/kT} - 1}. \tag{118.2}$$

Перейдем в этой формуле к пределу $\mathcal{E}_0 \to 0$. Тогда $e^{\mathcal{E}_0/kT} \approx$ $\approx 1 + \mathcal{E}_0/kT$, а потому $\bar{\mathcal{E}} = kT$. Получается классическое выражение для средней энергии осциллятора. Формула (118.2) в рассматриваемом предельном случае переходит, следовательно, в формулу Рэлея–Джинса. Этого и следовало ожидать, так как переход к пределу $\mathcal{E}_0 \to 0$ фактически означает возвращение к классическому рассмотрению, когда энергия осциллятора меняется не дискретными порциями, а непрерывно. Однако Планк, и это привело его к великому открытию, поставил вопрос, что получится, если предельный переход $\mathcal{E}_0 \to 0$ не производить. Величину $\mathcal{E}_0$ он определил из требования, чтобы выражение (118.2) удовлетворяло общей термодинамической формуле Вина

(116.10). Приравняв (118.2) и (116.10), видим, что это требование сводится к выполнению соотношения

$$\frac{1}{\pi^2 c^3}\frac{\mathcal{E}_0/\omega}{e^{\mathcal{E}_0/kT}-1} = f\left(\frac{\omega}{T}\right). \tag{118.3}$$

Но $\mathcal{E}_0$ есть характеристика *только самого осциллятора*, а потому не может зависеть от температуры T — макроскопического параметра, определяющего состояние вещества и излучения. Величина $\mathcal{E}_0$ может зависеть *только от собственной частоты ω осциллятора*. В таком случае, чтобы левая часть равенства (118.3) была функцией только аргумента ω/T, необходимо и достаточно, чтобы

$$\mathcal{E}_0 = \hbar\omega, \tag{118.4}$$

где $\hbar$ — постоянная. Она *универсальна*, поскольку в правой части (118.3) стоит универсальная функция $f(\omega/T)$ аргумента ω/T. Величина $\hbar$ называется *постоянной Планка*. Ее числовое значение по современным данным равно

$$\begin{aligned}\hbar &= (1{,}0545887 \pm 0{,}0000057)\cdot 10^{-27}\ \text{эрг}\cdot\text{с} = \\ &= (1{,}0545887 \pm 0{,}0000057)\cdot 10^{-34}\ \text{Дж}\cdot\text{с}.\end{aligned} \tag{118.5}$$

Сам Планк пользовался постоянной

$$\begin{aligned}h \equiv 2\pi\hbar &= (6{,}626176 \pm 0{,}000036)\cdot 10^{-27}\ \text{эрг}\cdot\text{с} = \\ &= (6{,}626176 \pm 0{,}000036)\cdot 10^{-34}\ \text{Дж}\cdot\text{с}.\end{aligned} \tag{118.5a}$$

Если теперь выражение (118.4) подставить в (118.2), то и получится формула Планка

$$u_\omega = \frac{\hbar\omega^3}{\pi^2 c^3}\frac{1}{e^{\hbar\omega/kT}-1}. \tag{118.6}$$

Формулу Планка обычно пишут в переменных ν, T:

$$u_\nu = \frac{8\pi h\nu^3}{c^3}\frac{1}{e^{h\nu/kT}-1}, \tag{118.7}$$

а также в переменных λ, T:

$$u_\lambda = \frac{8\pi hc}{\lambda^5}\frac{1}{e^{hc/\lambda kT}-1}. \tag{118.8}$$

При низких частотах, когда

$$\hbar\omega/kT \ll 1, \tag{118.9}$$

формула (118.6) переходит в формулу Рэлея–Джинса (117.8). В другом предельном случае высоких частот, когда

$$\hbar\omega/kT \gg 1, \tag{118.10}$$

получается формула

$$u_\omega = \frac{\hbar\omega^3}{\pi^2 c^3}\,e^{-\hbar\omega/kT}. \tag{118.11}$$

К формуле такого вида в 1896 г. пришел Вин на основе некоторых произвольных допущений.

Для высоких частот (ультрафиолет) формула Вина прекрасно согласуется с опытом. Однако в области низких частот (инфракрасная область спектра) она дает совершенно неверные результаты, здесь применима формула Рэлея–Джинса. Первоначально Планк и искал эмпирическую формулу, которая бы при низких частотах совпадала с формулой Рэлея–Джинса и непрерывно переходила в формулу Вина в области высоких частот.

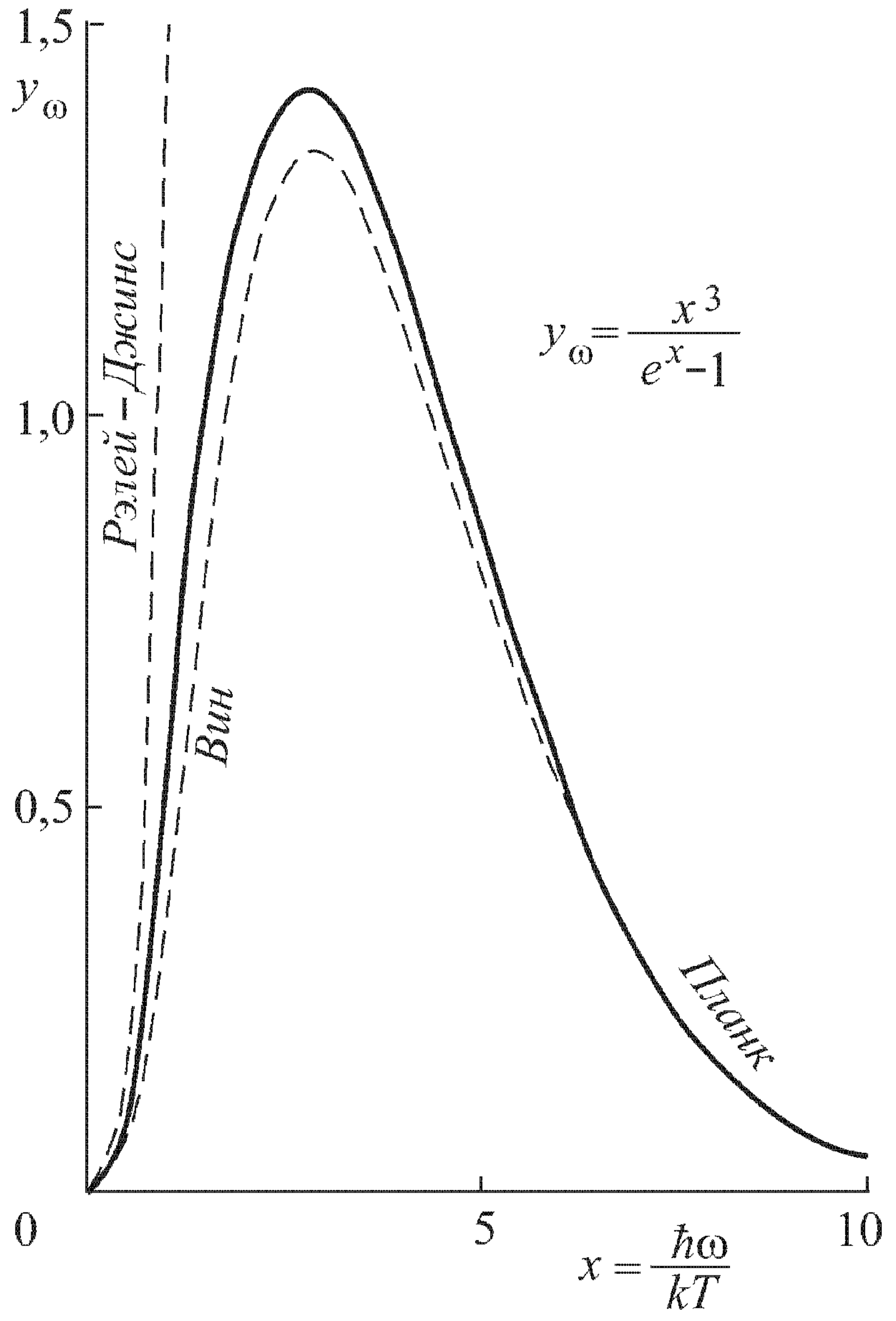


Рис. 343

Если ввести безразмерную переменную $x = \hbar\omega/kT$, то формулу (118.6) можно записать в виде

$$u_\omega = \frac{k^3T^3}{\pi^2c^3\hbar^2}\frac{x^3}{e^x-1}. \quad (118.6\text{а})$$

При введении другой безразмерной переменной $x = \lambda kT/hc$ получится

$$u_\lambda = \frac{8\pi k^5T^5}{h^4c^4}\frac{1/x^5}{e^{1/x}-1}. \quad (118.8\text{а})$$

Графики функций

$$y_\omega = \frac{x^3}{e^x-1} \quad \text{и} \quad y_\lambda = \frac{1/x^5}{e^{1/x-1}}$$

представлены на рисунках 343 и 344. Штриховые линии на тех же рисунках представляют те же функции, если их аппроксимировать по формулам Вина и Рэлея–Джинса.

3. Пользуясь формулой Планка, уточним значения постоянных в законах Стефана–Больцмана (115.4) и Вина (116.14), а также (116.15). Очевидно, что эти законы должны быть следствиями формулы Планка, так как последняя является частным случаем общей термодинамической формулы Вина (116.9). Согласно формуле Планка, интегральная плотность энергии равновесного излучения в вакууме равна

$$u = \frac{\hbar}{\pi^2c^3}\int\limits_0^\infty \frac{\omega^3\,d\omega}{\exp(\hbar\omega/kT)-1} = \frac{k^4T^4}{\pi^2c^3\hbar^3}\int\limits_0^\infty \frac{x^3e^{-x}\,dx}{1-e^{-x}}$$

(введена безразмерная переменная интегрирования $x = \hbar\omega/kT$).

Разложив знаменатель последнего интеграла $1 - e^{-x}$ в ряд и интегрируя, получим для последнего интеграла

$$\int\limits_0^\infty x^3 e^{-x}(1 + e^{-x} + e^{-2x} + \ldots)\,dx = 6\left(1 + \frac{1}{2^4} + \frac{1}{3^4} + \ldots\right) = \frac{\pi^4}{15},$$

так как по известной формуле сумма последнего ряда в скобках равна $\pi^4/90$. Таким образом,

$$u = \frac{\pi^2 k^4}{15 c^3 \hbar^3} T^4 = \frac{8}{15} \frac{\pi^5 k^4}{c^3 h^3} T^4, \tag{118.12}$$

т.е. получается закон Стефана–Больцмана (115.4), в котором постоянная a выражена через одни только фундаментальные постоянные c, h и k.

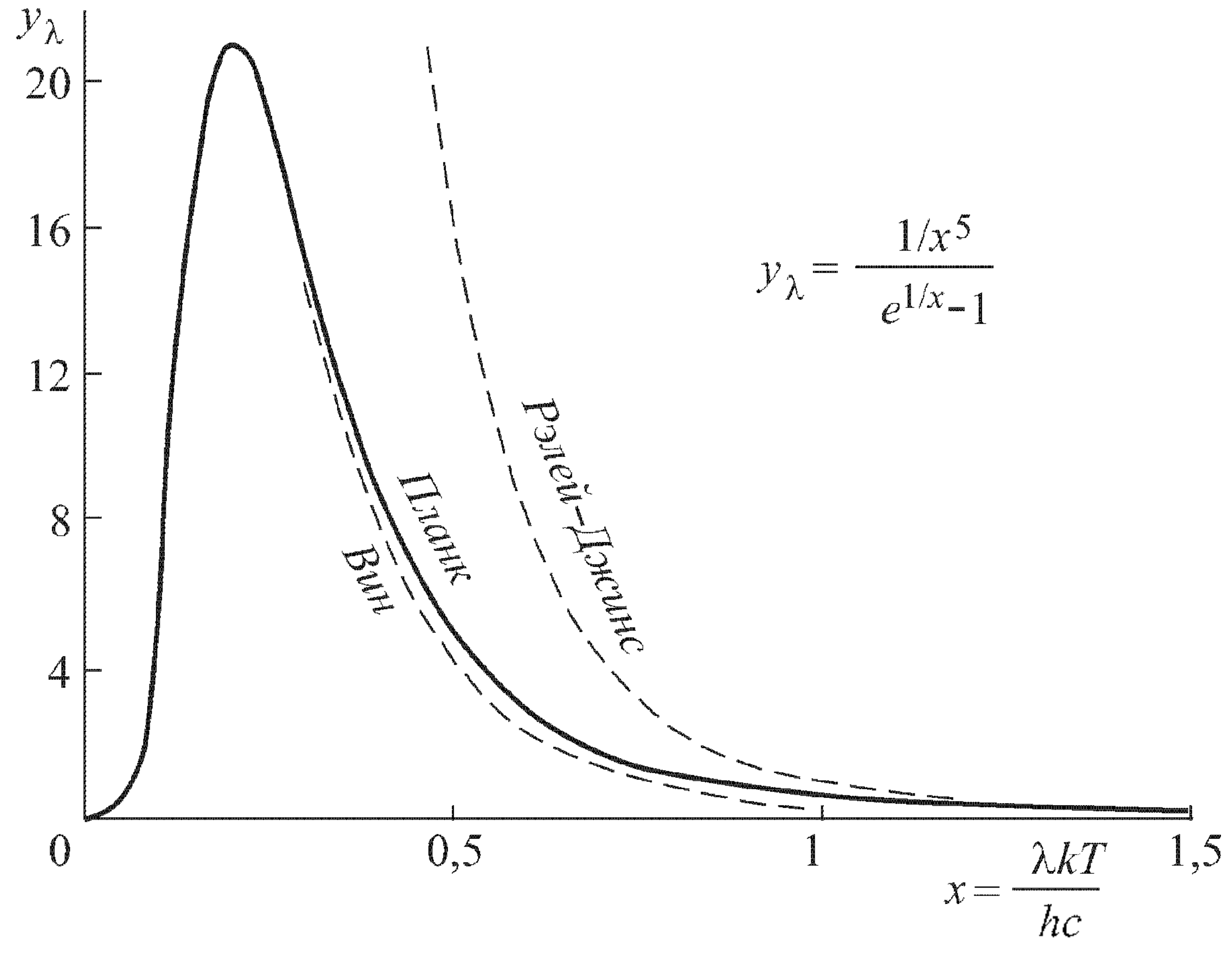


Рис. 344

На практике более удобна формула для энергетической светимости S излучающей абсолютно черной поверхности. Это есть интегральный лучистый поток, излучаемый наружу во всех направлениях (т.е. в телесный угол 2π) единицей площади такой поверхности в единицу времени. Она связана с яркостью B излучающей поверхности соотношением $S = \pi B = \pi I$ (см. § 22) или, ввиду формулы (112.5), $S = cu/4$. (Эта формула вполне аналогична выражению для среднего числа молекул газа, ударяющихся в единицу времени об единицу площади стенки сосуда, в который газ заключен, см. т. II, § 75.) Подставив сюда выражения (118.12), получим

$$S = \sigma T^4, \tag{118.13}$$

где

$$\sigma = \frac{\pi^2 k^4}{60c^2\hbar^3} = \frac{2\pi^5 k^4}{15c^2h^3} = (5{,}67032 \pm 0{,}00071) \cdot 10^{-8}\ \text{Вт}\cdot\text{м}^{-2}\cdot\text{К}^{-4}. \tag{118.14}$$

Величина σ называется *постоянной Стефана–Больцмана.*

4. Найдем теперь постоянную b в законе смещения Вина (116.14). Для этого надо найти значение $\lambda = \lambda_m$, для которого функция (118.8) при постоянном T обращается в максимум. Введем безразмерную переменную $\beta = hc/\lambda kT$ и выразим через нее функцию (118.8). Тогда, как легко убедиться, задача сводится к отысканию минимума функции $(e^\beta - 1)/\beta^5$. Приравняв нулю первую производную этой функции по β, получим уравнение

$$e^{-\beta} + \frac{\beta}{5} - 1 = 0, \tag{118.15}$$

корень которого $\beta = 4{,}9651142$. Поэтому

$$b = \lambda_m T = hc/(k\beta) = (2{,}897790 \pm 0{,}000090) \cdot 10^{-3}\ \text{м}\cdot\text{К}. \tag{118.16}$$

Если вместо λ пользоваться частотой ω, то закон Вина надо писать в виде (116.15). Тогда положение максимума, как нетрудно убедиться, будет определяться уравнением

$$(3 - \beta)e^\beta - 3 = 0, \tag{118.17}$$

где $\beta = \hbar\omega/kT = hc/(\lambda kT)$, т.е. β имеет тот же смысл, что и в предыдущем случае. Корень этого уравнения $\beta' = 2{,}8214393$. Если снова перейти к длинам волн, то максимуму функции u_ω соответствует длина волны λ'_m, определяемая условием

$$\lambda'_m T = hc/(k\beta'). \tag{118.18}$$

Таким образом, максимум на кривой частот сдвинут в длинноволновую сторону относительно максимума на кривой длин волн и притом так, что $\lambda'_m/\lambda_m = \beta/\beta' \approx 1{,}76$.

Заметим еще, что, измерив на опыте величины c, σ и b, можно по формулам (118.14) и (118.16) вычислить универсальные постоянные k и h, что и было впервые сделано Планком. После этого можно найти число Авогадро $N = R/k$ и элементарный заряд $e = F/N$, где R — универсальная газовая постоянная, а F — число Фарадея. Когда Планк производил эти вычисления, величины k, N и e были известны с малой точностью. Планк получил для них, а также для постоянной h значения, мало отличающиеся от современных.

5. Формулу Планка можно также получить, рассматривая равновесное излучение в полости как *фотонный газ*, к которому применима *статистика Бозе–Эйнштейна* (см. т. II, § 82).

Особенность этого газа состоит в том, что в результате взаимодействия с веществом фотоны могут рождаться и уничтожаться. Число их N в полости не остается постоянным. При равновесии оно устанавливается таким, что свободная энергия $F(T, V, N)$ при заданных T и V обращается в минимум, а потому $\partial F/\partial N = 0$. Но $\partial F/\partial N$ есть *химический потенциал* μ газа. Таким образом, для фотонов должно быть $\mu = 0$. Поэтому общая формула Бозе–Эйнштейна

$$\bar{n}_i = \frac{1}{\exp\left[(\mathcal{E}_i - \mu)/kT\right] - 1}, \tag{118.19}$$

определяющая среднее число частиц в i-м квантовом состоянии, в случае фотонов переходит в формулу

$$\bar{n} = \frac{1}{e^{\hbar\omega/kT} - 1}. \tag{118.20}$$

Число квантовых состояний в интервале частот $(\omega, \omega + d\omega)$ определяется выражением (117.7). Умножив $\bar{n}$ на это число, на энергию фотона $\hbar\omega$ и разделив на объем полости V, получим формулу Планка (118.6).

§ 119. Спонтанное и индуцированное излучение

1. В 1916 г. Эйнштейн дал новый вывод формулы Планка, основанный на представлениях Бора о механизме излучения. В этой работе было введено понятие *индуцированного излучения* — явления, на котором основан принцип действия лазера.

Пусть $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \ldots$ — значения энергии, которые может принимать атом или вообще любая атомная система. Атом может самопроизвольно перейти из высшего энергетического состояния $\mathcal{E}_n$ в низшее $\mathcal{E}_m$ с испусканием света. Такое излучение называется *спонтанным*. Если атом находится в световом поле, то последнее может вызывать переходы как с высшего уровня $\mathcal{E}_n$ на низший $\mathcal{E}_m$, так и обратно с низшего $\mathcal{E}_m$ на высший $\mathcal{E}_n$. Первые переходы сопровождаются излучением света. Оно и называется *индуцированным* (*вынужденным*) излучением. Обратные переходы сопровождаются поглощением света атомом.

Имеются аналоги описанных явлений и в классической физике. Если атом рассматривать как колебательную систему, то в поле световой волны она будет совершать вынужденные колебания. В зависимости от соотношения фаз между колебаниями этой системы и светового поля амплитуда колебаний атома может как увеличиваться (поглощение света), так и уменьшаться (вынужденное излучение).

Эйнштейн применил к описанию процессов спонтанного и вынужденного излучения *вероятностные методы*. При этом для проблемы равновесного излучения не имеет значения, присуща ли вероятность ансамблю физических объектов или самим элементарным законам, управляющим их поведением.

Рассмотрим теперь много одинаковых атомов в световом поле. Последнее будем предполагать изотропным и неполяризованным. Тогда отпадает вопрос о зависимости коэффициентов, вводимых ниже, от поляризации и направления излучения. Пусть N_m и N_n — числа атомов в состояниях $\mathcal{E}_m$ и $\mathcal{E}_n$, причем состояния $\mathcal{E}_m$ и $\mathcal{E}_n$ могут быть взяты какими угодно из ряда допустимых состояний $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, ... Среднее число переходов атомов из состояния $\mathcal{E}_n$ в состояние $\mathcal{E}_m$ в единицу времени из-за спонтанного излучения будет пропорционально исходному числу атомов N_n. Представим его в виде $A_n^m N_n$. Эйнштейн постулировал, что из-за индуцированного излучения среднее число переходов между теми же уровнями будет по-прежнему пропорционально N_n, а также спектральной плотности излучения $u(\omega_{mn})$ при частоте испускаемого света, соответствующей рассматриваемому переходу. Обозначим это число через $B_n^m N_n u(\omega_{mn})$. Аналогично, среднее число переходов с уровня $\mathcal{E}_m$ на уровень $\mathcal{E}_n$ из-за поглощения света представится как $B_m^n N_m u(\omega_{mn})$. Величины A_n^m, B_n^m, B_m^n называются *коэффициентами Эйнштейна*. Они являются характеристиками *только самого атома* и могут зависеть лишь от частоты ω_{mn}.

2. Допустим теперь, что поле излучения, в котором находятся атомы, равновесное и имеет температуру T. Тогда имеет место детальное равновесие, а потому

$$A_n^m N_n + B_n^m N_n u(\omega_{mn}) = B_m^n N_m u(\omega_{mn}). \tag{119.1}$$

Если уровни энергии $\mathcal{E}_m$ и $\mathcal{E}_n$ *простые*, а не кратные, то коэффициенты Эйнштейна связаны соотношением

$$B_m^n = B_n^m. \tag{119.2}$$

Действительно, будем повышать температуру системы. Коэффициенты Эйнштейна при этом меняться не будут, так как они от температуры не зависят. Спонтанное же излучение будет играть все меньшую и меньшую роль по сравнению с индуцированным. Если им пренебречь, то условие детального равновесия примет вид $B_n^m N_n = B_m^n N_m$. Но, согласно формуле Больцмана, при $T \to \infty$ населенности уровней N_n и N_m должны сравняться. Отсюда и следует, что $B_m^n = B_n^m$.

Допустим теперь, что уровень $\mathcal{E}_m$ состоит из g_m, а уровень $\mathcal{E}_n$ — из g_n слившихся простых энергетических подуровней. Такие уровни называются *кратными*, а целые числа g_m и g_n — их *кратностями*. Вероятность перехода атома с уровня $\mathcal{E}_m$ на каждый простой подуровень $\mathcal{E}_n$ меньше вероятности перехода на кратный уровень $\mathcal{E}_n$ в g_n раз, т.е. она равна uB_m^n/g_n. Аналогично, вероятность перехода с уровня $\mathcal{E}_n$ на простой подуровень $\mathcal{E}_m$

будет uB_n^m/g_m. Но по доказанному эти вероятности равны между собой. Поэтому для кратных уровней соотношение (119.2) заменится на $B_m^n/g_n = B_n^m/g_m$ или

$$g_m B_m^n = g_n B_n^m. \tag{119.3}$$

Так как спектральная плотность излучения $u(\omega)$ не зависит от того, какие атомы использованы для ее вычисления, то для простоты расчета энергетические уровни $\mathcal{E}_n$, $\mathcal{E}_m$ можно считать простыми. Тогда с учетом формулы Больцмана

$$\frac{N_m}{N_n} = \exp\frac{\mathcal{E}_n - \mathcal{E}_m}{kT}$$

из (119.1) получим

$$u(\omega_{mn}) = \frac{\alpha(\omega_{mn})}{\exp\left[(\mathcal{E}_n - \mathcal{E}_m)/(kT)\right] - 1}, \tag{119.4}$$

где $\alpha(\omega_{mn}) = A_n^m/B_n^m$. Сравнение этого выражения с формулой Вина (116.10) показывает, что разность $\mathcal{E}_n - \mathcal{E}_m$ должна быть линейной функцией частоты, т.е.

$$\mathcal{E}_n - \mathcal{E}_m = \hbar\omega_{mn}. \tag{119.5}$$

В предельном случае низких частот (или высоких температур)

$$u(\omega_{mn}) = \frac{\alpha(\omega_{mn})kT}{\hbar\omega_{mn}}.$$

Но в этом случае применима формула Рэлея–Джинса, так что

$$\alpha(\omega_{mn}) = \frac{\hbar\omega_{mn}^3}{\pi^2 c^3}. \tag{119.6}$$

Подставим теперь выражения (119.5) и (119.6) в формулу (119.4). При этом индексы m и n можно опустить, так как в нашем выводе уровни энергии $\mathcal{E}_m$ и $\mathcal{E}_n$ можно взять произвольными. Тогда получится формула Планка (118.6).

Из универсальности функции (118.6) следует, что и постоянная $\hbar$, введенная посредством соотношения (119.5), также универсальна. Формула (119.5), определяющая частоту излучаемого света при квантовых переходах между энергетическими уровнями, называется *правилом частот Бора*.

Заметим еще, что если бы не учитывать индуцированное излучение, т.е. положить $B_n^m u(\omega_{mn}) = 0$, то вместо формулы Планка получился бы ее предельный случай — формула Вина (118.11). Отсюда следует, что *формула Планка с неизбежностью приводит к заключению о существовании индуцированного излучения.* При низких температурах индуцированное излучение, очевидно, несущественно по сравнению со спонтанным. Вот почему в области низких температур, когда $\hbar\omega/kT \gg 1$, справедлива формула Вина.

ЗАДАЧИ

1. Определить среднее число фотонов n в единице объема полости, заполненной равновесным (черным) излучением, при температуре T.

Ответ.

$$n = \frac{k^3T^3}{\hbar^3\pi^2c^3}\int\limits_0^\infty \frac{x^2e^{-x}}{1-e^{-x}}\,dx =$$

$$= \frac{2k^3T^3}{\hbar^3\pi^2c^3}\left(1+\frac{1}{2^3}+\frac{1}{3^3}+\ldots\right) \approx 1{,}202\frac{2k^3T^3}{\hbar^3\pi^2c^3} = 20{,}5T^3,$$

где T — температура в кельвинах.

2. В полости с зеркальными стенками находится изотропное излучение с средней плотностью n фотонов в единице объема. Определить среднее число фотонов z, ежесекундно ударяющихся об единицу площади стенки.

Ответ. $z = \dfrac{1}{4}nc$.

3. Решить предыдущую задачу в предположении, что излучение равновесное, а температура стенок полости равна T. Какое среднее число фотонов N будет выходить ежесекундно из полости через отверстие в стенке полости площадью $s = 1$ см^2, если $T = 1000$ К?

Ответ. $n \approx 2{,}404\dfrac{k^3T^3}{\hbar^3\pi^2c^3}$, $N = 0{,}6\dfrac{k^3T^3}{\hbar^3\pi^2c^2} \approx 1{,}5\cdot 10^{20}$.

4. В какой области спектра равновесного (черного) излучения при температуре $T = 300$ К интенсивность индуцированного излучения превосходит интенсивность спонтанного?

Ответ. Если $\lambda \geqslant hc/(kT\ln 2)$, то интенсивность индуцированного излучения становится равной или большей интенсивности спонтанного. При $T = 300$ К получаем $\lambda \geqslant 692$ мкм.

5. При какой температуре равновесного (черного) излучения индуцированное излучение в видимой области ($\lambda = 550$ нм) превосходит спонтанное?

Ответ. $T > hc/(k\lambda\ln 2) \approx 3{,}8\cdot 10^4$ К.

6. Среднее число фотонов в единице объема равновесного (черного) излучения, приходящееся на интервал частот $(\omega, \omega + d\omega)$ или на соответствующий ему интервал длин волн $(\lambda, \lambda + d\lambda)$, можно представить в виде $dN = f(\omega, T)\,d\omega = \varphi(\lambda, T)\,d\lambda$, где T — температура излучения. Найти положение максимумов функций $f(\omega, T)$ и $\varphi(\omega, T)$ при фиксированной температуре T.

Ответ. Для функции $f(\omega, T)$ максимум получается при $\hbar\omega/(kT) = x$, где x — корень уравнения $(x-2)e^x + 2 = 0$, т.е. $x = 1{,}593624$. Соответствующая длина волны λ_1 находится из соотношения

$$\lambda_1 T = \frac{2\pi\hbar c}{kx} = \frac{hc}{kx} = 0{,}902867 \text{ см·К}.$$

Для функции $\varphi(\lambda, T)$ максимум получается при $\lambda = \lambda_2$, причем $\dfrac{hc}{\lambda kT} = \xi$, где ξ — корень уравнения $(\xi - 4)e^\xi + 4 = 0$, т.е. $\xi = 3{,}920690$. Таким образом, $\lambda_2 T = 0{,}368967$ см·К. Следовательно, $\lambda_2 < \lambda_1$.

7. Равновесное излучение заключено в полости, стенки которой поддерживаются при постоянной (абсолютной) температуре T. Вычислить флуктуации энергии $\mathcal{E}$ такого излучения в объеме V в спектральном интервале

$(\omega, \omega + d\omega)$, пользуясь формулами 1) Вина, 2) Рэлея–Джинса, 3) Планка, и интерпретировать полученные результаты с точки зрения корпускулярных и волновых представлений о свете.

Решение. Применяя к (97.15) формулы Вина, Рэлея–Джинса и Планка, получим

$$\overline{\Delta\mathcal{E}^2} = \bar{\mathcal{E}}\hbar\omega \qquad \text{(Вин)}, \tag{119.7}$$

$$\overline{\Delta\mathcal{E}^2} = \bar{\mathcal{E}}kT = \frac{\pi^2 c^3}{V\omega^2\, d\omega}\,\overline{\mathcal{E}}^2 \qquad \text{(Рэлей- Джинс)}, \tag{119.8}$$

$$\overline{\Delta\mathcal{E}^2} = \bar{\mathcal{E}}\hbar\omega + \frac{\pi^2 c^3}{V\omega^2\, d\omega}\,\overline{\mathcal{E}}^2 \qquad \text{(Планк)}. \tag{119.9}$$

Формула (119.7) имеет такой же вид, что и формула (97.7) для флуктуации числа частиц идеального газа. Ее можно было бы получить из корпускулярных представлений, рассматривая излучение как газ независимых частиц. Напротив, формула (119.8) соответствует волновым представлениям о свете. Здесь флуктуации возникают из-за суперпозиции волн различных частот. Формула (119.9) соответствует синтезу обоих представлений.

8. Освещенность E, создаваемая звездой первой величины на поверхности Земли при нормальном падении света, составляет около 10^{-6} лк. Можно ли объяснить мерцание звезд квантовыми флуктуациями света?

Решение. Среднее число фотонов, попадающее от звезды в зрачок глаза в одну секунду,

$$\bar{N} = \frac{EAs}{h\lambda/c} \approx 6 \cdot 10^{34} \sim 10^{34},$$

где A — механический эквивалент света, площадь зрачка глаза s принята равной 0,5 см^2, а длина волны $\lambda = 550$ нм. В рассматриваемой области спектра с хорошим приближением можно пользоваться формулой Вина, а потому

$$\overline{\Delta N^2} = \bar{N}, \qquad \frac{1}{\bar{N}}\sqrt{\overline{\Delta N^2}} \sim 10^{-17}.$$

Отсюда видно, что квантовые флуктуации света к мерцанию звезд не имеют отношения.

ГЛАВА XI

ЛАЗЕРЫ И НЕЛИНЕЙНАЯ ОПТИКА

§ 120. Принципы работы лазера

1. Спонтанное излучение некогерентно. В этом случае атомы источника излучают свет независимо друг от друга. Фазы волн, испускаемых различными атомами, их поляризация и направления распространения никак не связаны между собой. Обычные источники света — пламена, лампы накаливания, газоразрядные трубки, люминесцентные лампы и пр. — излучают некогерентно. В них свечение вызывается либо столкновениями между атомами, совершающими тепловое движение, либо электронными ударами. Правда, в таких источниках наряду со спонтанным происходит и индуцированное излучение. Однако оно возбуждается некогерентным спонтанным излучением, а потому и само некогерентно. Испускаемый свет характеризуется большей или меньшей степенью беспорядка. Максимальный беспорядок достигается в равновесном излучении в полости. В нем представлены всевозможные фазы и частоты, всевозможные направления колебаний, всевозможные направления распространения света. Если заимствовать терминологию из акустики и радиотехники, то можно сказать, что указанные источники света генерируют не правильные или упорядоченные волны, а шумы, пригодные только для освещения, грубой сигнализации, получения изображений, фотографирования и пр., но не для передачи речи, телевидения и т.д., осуществляющихся посредством радиоволн, излучаемых радиостанциями.

Однако можно создать и когерентно излучающие источники света, в которых бы различные атомы излучали волны согласованно, подобно радиостанциям, т.е. с одинаковыми частотами, фазами, поляризацией и направлением распространения. Такие источники открыли широкие возможности для разнообразных научных и технических применений. Они называются *оптическими квантовыми генераторами* или *лазерами.* Слово «лазер» образовалось из первых букв полного английского названия «Light amplification by stimulated emission of radiation», что в переводе означает: усиление света посредством индуцированного излучения. Созданию лазеров предшествовало изобретение *мазеров*, т.е. усилителей микроволн, работающих также на принципе индуцированного излучения. Поэтому первоначально лазеры назывались *оптическими мазерами*. Подробное рассмотрение устройства и работы лазеров и мазеров дается в квантовой электронике. Здесь, в оптике, мы ограничимся только кратким изложением принципов работы лазеров.

Лазер работает на принципе *индуцированного излучения*. Допустим, что на атом падает фотон с энергией $\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1$, где $\mathcal{E}_1$ и $\mathcal{E}_2$ — какие-либо два энергетических уровня атома. Если атом находится на нижнем уровне $\mathcal{E}_1$, то падающий фотон может поглотиться. Если же атом находится на верхнем уровне $\mathcal{E}_2$, то может произойти вынужденный переход на нижний уровень $\mathcal{E}_1$ с испусканием второго фотона. Индуцированно излученный фотон характеризуется не только *той же частотой* ω (как и при спонтанном излучении), но также *теми же фазой*, *поляризацией* и *направлением распространения*. Вместо одного падающего фотона получается *два тождественных фотона*. Эта особенность индуцированного излучения и используется в лазерах.

2. Рассмотрим теперь не единичный атом, а среду из атомов. Обозначим через N_1 и N_2 числа атомов в единице объема на уровнях $\mathcal{E}_1$ и $\mathcal{E}_2$ соответственно. Допустим, что в среде распространяется плоская монохроматическая волна, частота которой определяется условием $\hbar\omega = \mathcal{E}_2 - \mathcal{E}_1$. За время $dt = dx/v$, где v — скорость распространения, а dx — расстояние, пройденное волной, с нижнего уровня на верхний переходит в среднем $u(\omega)B_1^2 N_1\,dt$ атомов и такое же число фотонов поглощается. Из-за индуцированного излучения с верхнего уровня на нижний перейдет $u(\omega)B_2^1 N_2\,dt$ атомов и родится такое же число фотонов той же поляризации и направления распространения, что и у рассматриваемой волны. Фотоны, излученные спонтанно, а также фотоны, индуцированные другими волнами, можно не учитывать, так как среди них только ничтожная часть распространяется в нужном направлении и обладает нужной поляризацией. Увеличение числа фотонов в единице объема при прохождении волной расстояния $dx = v\,dt$ представится выражением

$$dN_{\text{фот}} = (B_2^1 N_2 - B_1^2 N_1)u(\omega)\frac{dx}{v}.$$

Коэффициенты Эйнштейна B_2^1 и B_1^2 связаны соотношением $g_2 B_2^1 = g_1 B_1^2$, где g_1 и g_2 — кратности уровней $\mathcal{E}_1$ и $\mathcal{E}_2$. Используя это соотношение, перепишем предыдущее уравнение в виде

$$dN_{\text{фот}} = \left(\frac{N_2}{g_2} - \frac{N_1}{g_1}\right) g_1 B_1^2 u(\omega)\frac{dx}{v}. \tag{120.1}$$

Чтобы при распространении в среде волна усиливалась, необходимо выполнение условия

$$\frac{N_2}{g_2} > \frac{N_1}{g_1}. \tag{120.2}$$

Его можно записать в виде

$$n_2 > n_1, \tag{120.3}$$

где $n_1 = N_1/g_1$ и $n_2 = N_2/g_2$ — числа атомов на каждом из простых уровней, из которых состоят сложные уровни $\mathcal{E}_1$ и $\mathcal{E}_2$.

В обычных условиях, когда среда находится в *термодинамическом равновесии*, $n_2 < n_1$, т.е. *на каждом простом верхнем уровне находится меньше атомов, чем на нижнем.* Это непосредственно следует из формулы Больцмана

$$n = n_0 e^{-\mathcal{E}/kT}.$$

Можно искусственно получить *термодинамически неравновесную среду*, у которой выполняется соотношение, обратное (120.3) или (120.2). Такая среда называется *активной* или *средой с инверсной заселенностью по отношению к энергетическим уровням* $\mathcal{E}_1$ *и* $\mathcal{E}_2$. Следовательно, *для усиления световой волны необходимо, чтобы среда, в которой волна распространяется, была активной.* Идея использования индуцированного излучения для усиления волны была впервые высказана в 1939 г. в докторской диссертации В.А. Фабрикантом (р. 1907) и впоследствии (в 1951 г.) на нее было выдано авторское свидетельство. В то время на идею Фабриканта не было обращено должного внимания. Казалось, что создание систем с инверсной заселенностью энергетических уровней — дело бесперспективное.

3. Усиление света в активной среде обычно сравнивают с нарастанием лавины, изображая фотоны в виде шариков. Летящий фотон-шарик порождает второй фотон-шарик с переходом атома с верхнего уровня на нижний. Получаются два одинаковых шарика, летящих в прежнем направлении, затем четыре шарика и т.д. Но эта грубая иллюстрация не объясняет, как в результате наложения фотонов формируется монохроматическая волна строго определенного направления. Эта сторона дела становится понятной, если сравнить изучаемое нами явление с классической картиной распространения плоской монохроматической волны в однородной среде. Волна вызывает колебания в атомах и молекулах среды. Последние переизлучают шаровые волны, когерентные друг с другом и с падающей волной. Эти шаровые волны, интерферируя между собой, создают снова плоский волновой фронт, распространяющийся в среде. Они влияют только на фазовую скорость волны. Если среда абсолютно прозрачна, то амплитуда волны должна оставаться постоянной, как того требует закон сохранения энергии. В поглощающих средах энергия волны частично переходит в тепло — амплитуда волны убывает. Но в активной среде молекулы и атомы находятся в возбужденных состояниях. За счет энергии возбуждения вторичные световые волны, излучаемые молекулами и атомами, усиливаются. Однако их фазы и поляризация остаются прежними. Поэтому остаются прежними поляризация и фаза также и результирующей волны, возникающей в результате интерференции таких вторичных волн. Усиливается только ее амплитуда.

4. Индуцированное излучение было использовано для *генерации когерентных световых волн.* Идея этого впервые была высказана в 1957 г. А.М. Прохоровым (р. 1916) и Н.Г. Басовым (р. 1922) и независимо от них Ч. Таунсом (р. 1915). Чтобы активное вещество превратить в генератор световых колебаний, надо осуществить *обратную связь.* Необходимо, чтобы часть излученного света все время находилась в зоне активного вещества и вызывала вынужденное излучение все новых и новых атомов. Для этого активное вещество помещают между двумя параллельными зеркалами. Допустим, например, что оно представляет собой цилиндр, а плоскости зеркал S_1 и S_2 перпендикулярны к оси этого цилиндра (рис. 345). Тогда луч света, претерпевая многократные отражения от зеркал S_1 и S_2, будет проходить много раз через активное вещество, усиливаясь при этом в результате вынужденных переходов атомов с высшего энергетического уровня $\mathcal{E}_2$ на более низкий уровень $\mathcal{E}_1$. Получается *открытый резонатор*, представляющий собой в сущности интерферометр Фабри–Перо, только заполненный активной средой.

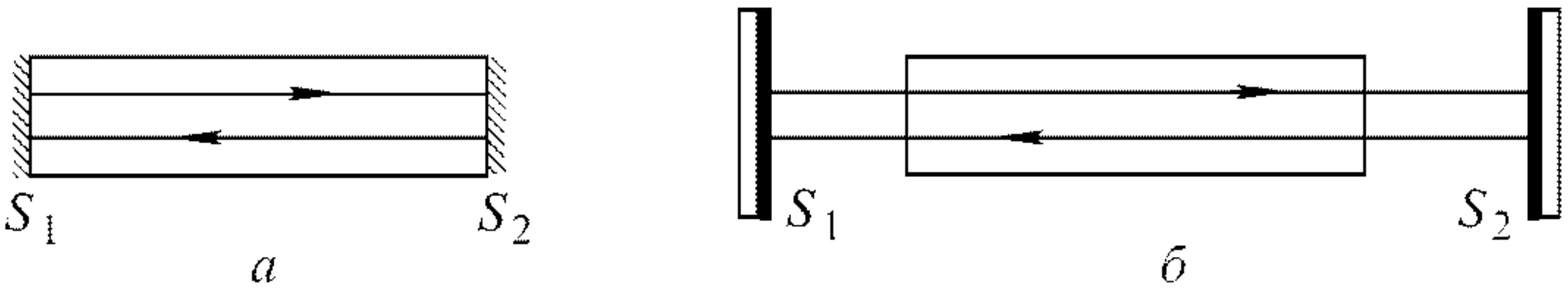


Рис. 345

Такой резонатор будет *не только усиливать свет*, но также *коллимировать* и *монохроматизировать* его. Для простоты предположим сначала, что зеркала S_1 и S_2 идеальные. Тогда лучи, параллельные оси цилиндра, будут проходить через активное вещество туда и обратно неограниченное число раз. Все же лучи, идущие наклонно, в конце концов попадут на боковую стенку цилиндра, где они рассеются или выйдут наружу. Ясно поэтому, что максимально усилятся лучи, распространяющиеся параллельно оси цилиндра. Этим и объясняется коллимация лучей. Конечно, строго параллельные лучи получить нельзя. Этому препятствует дифракция света. Угол расхождения лучей принципиально не может быть меньше *дифракционного предела* $\delta\vartheta \approx \lambda/D$, где D — ширина пучка. Однако в лучших газовых лазерах такой предел практически достигнут.

5. Объясним теперь, как происходит *монохроматизация света*. Для простоты проведем рассуждение применительно к рис. 345 *а*, когда роль зеркал S_1 и S_2 выполняют отполированные и посеребренные торцы цилиндра активного вещества, перпендикулярные к его геометрической оси. Пусть L — длина цилиндра. Если $2L = m\lambda$, т.е. на длине L укладывается целое число полуволн m, то световая волна, выйдя от S_1, после прохождения через цилиндр туда и обратно вернется к S_1 в той же фазе. Такая волна усилится при втором и всех следующих прохождениях через активное вещество в прямом и обратном направлениях. Ближайшая длина волны $\lambda \pm \Delta\lambda$, для которой должно происходить такое же усиление, найдется из условия $2L = (m \pm 1)(\lambda \mp \Delta\lambda)$. Следовательно, $\Delta\lambda = \lambda/m$, т.е. $\Delta\lambda$, как и следовало ожидать, совпадает со спектральной областью интерферометра Фабри–Перо.

Учтем теперь, что энергетические уровни $\mathcal{E}_1$ и $\mathcal{E}_2$ и спектральные линии, возникающие при переходах между ними, не бесконечно тонкие, а имеют *конечную ширину*. Предположим сначала, что ширина спектральной линии, излучаемой атомами, меньше дисперсионной области прибора. Тогда из всех длин волн, излучаемых атомами, условию $2L = m\lambda$ может удовлетворять только одна длина волны λ. Такая волна усилится максимально. Это и ведет к сужению спектральных линий, генерируемых лазером, т.е. к монохроматизации света.

Степень монохроматизации нетрудно определить. Пусть свет проходит через активное вещество туда и обратно N раз. Для длины волны λ имеем $2LN = Nm\lambda$. Возьмем ближайшую длину волны λ', удовлетворяющую условию $2LN = (Nm \pm 1)\lambda'$. Для такой длины волны каждый цуг волн, возникший при прохождении через активное вещество туда и обратно, будет отличаться по фазе от предыдущего и последующего цугов на $\pm 2\pi/N$. В результате все N цугов погасят друг друга — получится минимум интенсивности для λ'. Отсюда ясно, что ширина спектральной линии, усиливаемой лазером, будет $\delta\lambda = |\lambda - \lambda'| \sim \lambda/(Nm)$, т.е. она определяется разрешающей силой прибора. При $N \to \infty$ получится $\delta\lambda \to 0$, т.е. бесконечно тонкая спектральная линия. В действительности из-за неидеальности отражающих поверхностей зеркал $\delta\lambda$ остается конечной. Однако при хороших отражающих поверхностях лазер дает очень тонкую, практически монохроматическую линию. Допустим теперь, что спектральные линии, излучаемые атомами активной среды, шире дисперсионной области прибора $\Delta\lambda$. В этом случае вместо усиления одной спектральной линии может возникнуть усиление нескольких линий. Свет лазера будет состоять из близких, практически монохроматических линий.

6. В приведенных рассуждениях не принято во внимание, что в резонаторе Фабри–Перо могут усиливаться волны, распространяющиеся не только параллельно его оси, но и *под малыми углами* φ к ней. Условие усиления имеет вид $2L\cos\varphi = m\lambda$. Это, как и во всяком объемном резонаторе, создает систему стоячих волн с узлами и пучностями. Каждой такой волне соответствует определенный тип или, как принято говорить, *мода колебаний*. Это обстоятельство отражается на распределении интенсивности в поперечном сечении пучка лучей, генерируемых лазером. Здесь наблюдается система светлых пятен, разделенных темными узловыми линиями. Это, конечно, усложняет дело. Но мы не будем входить в подробное обсуждение этого вопроса. Заметим только, что наряду с плоскими зеркалами в резонаторах употребляются длиннофокусные *сферические зеркала*. Они могут быть, например, вогнутыми конфокальными зеркалами, т.е. зеркалами, у которых главные фокусы совпадают и находятся в середине системы. Применение подобных зеркал, как показывают опыт и численные расчеты, позволяет значительно уменьшить потери света в лазерах и упрощает их юстировку.

7. Учтем теперь, что в реальном лазере часть света, чтобы ее можно было использовать, должна быть выпущена из активной среды наружу. С этой целью одно из зеркал, например S_2, делается *полупрозрачным*. Кроме того, и зеркало S_1 лишь частично отражает свет, хотя коэффициент отражения его и близок к 100 %. Это приводит к ослаблению светового пучка. Чтобы лазер был генератором света, необходимо, чтобы усиление светового пучка в активной среде превосходило некоторое минимальное — *пороговое* — значение. Именно, должно быть выполнено следующее условие. Световой пучок, вышедший от S_1, после прохождения туда и обратно через активную среду и отражения от S_1 должен вернуться в исходное положение с неменьшей интенсивностью. Иначе в результате последовательного повторения этих процессов интенсивность пучка будет непрерывно убывать и лазер перестанет генерировать. Поэтому для генерации недостаточно выполнения простого неравенства (120.3). Оно должно быть выполнено с некоторым запасом, т.е. число атомов N_2 на верхнем уровне в единице объема активной среды должно превышать некоторое минимальное — пороговое — значение.

Конечно, нарастание интенсивности волны в активной среде не может продолжаться беспредельно, так как заселенность верхнего энергетического уровня ограничена. По мере обеднения атомами верхнего уровня $\mathcal{E}_2$ скорость нарастания интенсивности волны будет уменьшаться и волна начнет затухать еще до того, как перестанет выполняться условие (120.3).

§ 121. Рубиновый лазер

1. Существует много материалов, применяемых в качестве активных сред в лазерах. Сюда относятся различные диэлектрические кристаллы, стекла, газы, полупроводники и даже плазма. Эта сторона дела подробно изучается в квантовой электронике. Мы ограничимся описанием устройства и принципа действия только двух лазеров: *рубинового* и *гелий-неонового.*

Первый квантовый генератор света был создан в 1961 г. Мейманом (р. 1927) на рубине. Рубин — это твердый кристалл, основой которого является корунд, т.е. кристалл окиси алюминия (Al_2O_3), в котором небольшая часть атомов алюминия (около 0,05 %) замещена ионами хрома Cr^{+++} [1]. Ионы хрома играют основную роль в работе квантового генератора. Корунд — это диэлектрик с широкой запрещенной энергетической зоной между валентной зоной и зоной проводимости (см. т. III, § 100). Энергетические уровни хрома в корунде лежат в этой запрещенной зоне. Они были изучены задолго до создания лазеров. Основным (невозбужденным) уровнем является уровень $\mathcal{E}_1$ (рис. 346). Он имеет сложную структуру, которая, однако, не играет роли в работе квантового генератора. Выше лежат близкие возбужденные уровни $\mathcal{E}_{2а}$ и $\mathcal{E}_{2б}$. Это — узкие уровни. При переходе с них на основной уровень $\mathcal{E}_1$ излучается красный свет с длинами волн $\lambda = 694{,}3$ и 692,9 нм. Он-то и придает рубину характерную розовую или

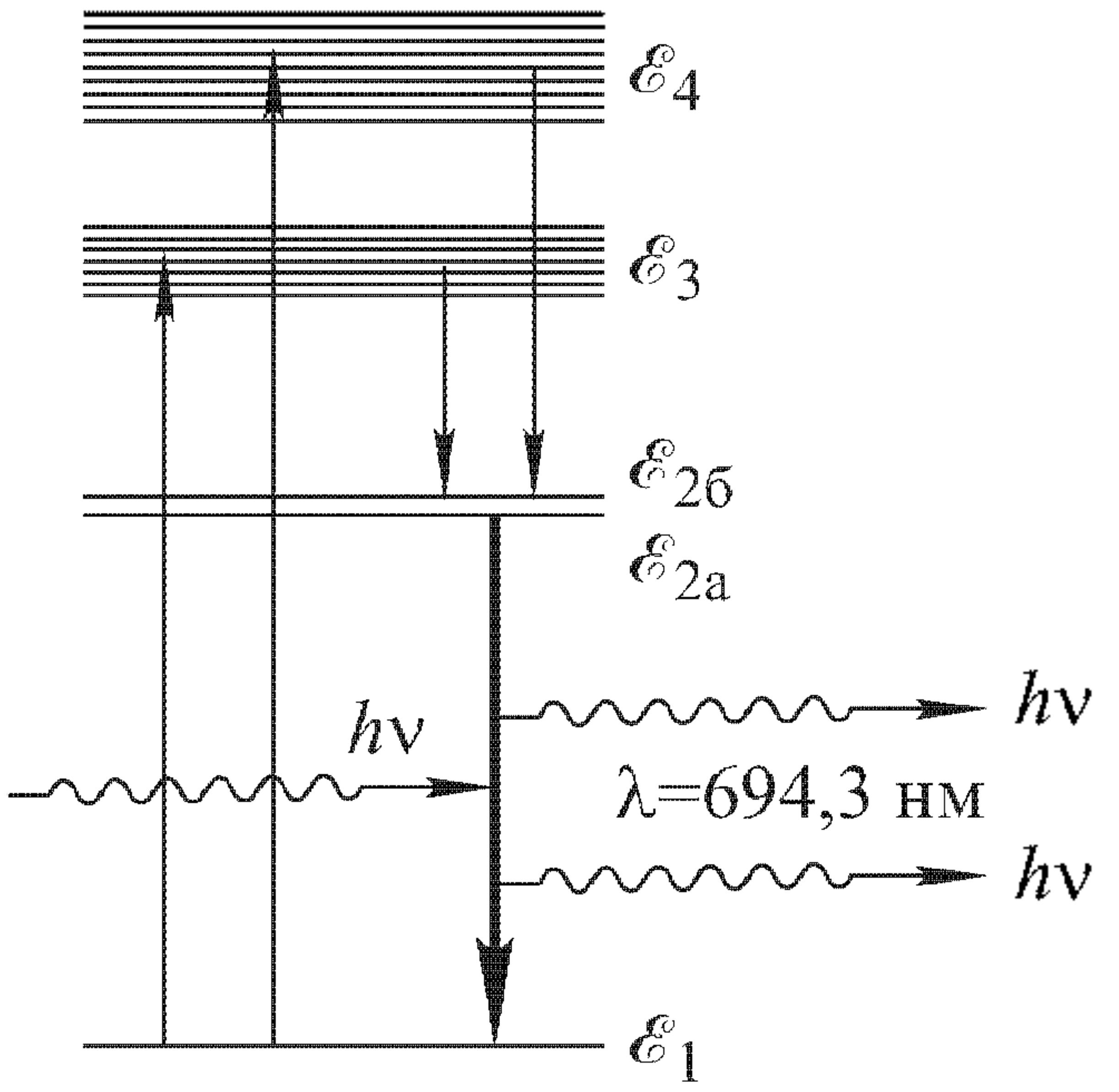


Рис. 346

---

[1] Для лазеров монокристаллы рубина выращивают искусственно. Смесь Al_2O_3 и Cr_2O_3 в виде пудры сыплется сверху на выращиваемый кристалл рубина, верхняя кромка которого находится в кислородно-водородном пламени горелки с температурой 2050 °C, достаточной для плавления рубина. При медленном опускании кристалла расплавленный слой смеси выходит из пламени и кристаллизуется. Таким путем удается получать монокристаллы рубина в виде стержней длиной до 0,5 м и диаметром до 5 см.

красную окраску (в зависимости от концентрации ионов хрома). Более интенсивна линия $\lambda = 694{,}3$ нм. По этой причине только она и усиливается при работе лазера. Вторая линия не играет роли. Выше уровней $\mathcal{E}_{2\text{а}}$ и $\mathcal{E}_{2\text{б}}$ расположены две сравнительно широкие полосы энергий $\mathcal{E}_3$ и $\mathcal{E}_4$. Переходы между уровнями этих полос и основным уровнем $\mathcal{E}_1$ сопровождаются излучением зеленого и голубого света соответственно.

2. Инверсная заселенность создается между уровнями $\mathcal{E}_1$ и $\mathcal{E}_2$. Для этого используется так называемая *оптическая накачка*, т.е. освещение кристалла рубина мощной вспышкой света. Рубину придают форму цилиндрического стержня с диаметром 0,1–2 см и длиной от 2 до 20 см и больше. Концы стержня тщательно отполированы. Они могут служить зеркалами. Тогда их серебрят, как показано на рис. 345 *а*. Зеркала могут быть и внешними (рис. 345 *б*), тогда серебрение не нужно. Для освещения рубинового стержня применяют импульсные ксеноновые газоразрядные *лампы-вспышки*, через которые разряжаются батареи высоковольтных конденсаторов (напряжение 2–3 кВ). Длительность вспышки порядка одной миллисекунды. Лампа-вспышка имеет форму спиральной трубки, обвивающейся вокруг рубинового стержня (рис. 347). Она может быть и прямолинейной. Тогда применяют зеркальные осветители, имеющие форму эллиптических цилиндров с внутренними отражающими поверхностями. Лампа-вспышка располагается вдоль одной из фокальных линий цилиндра; отраженный свет концентрируется на рубиновом стержне, помещаемом вдоль другой фокальной линии.

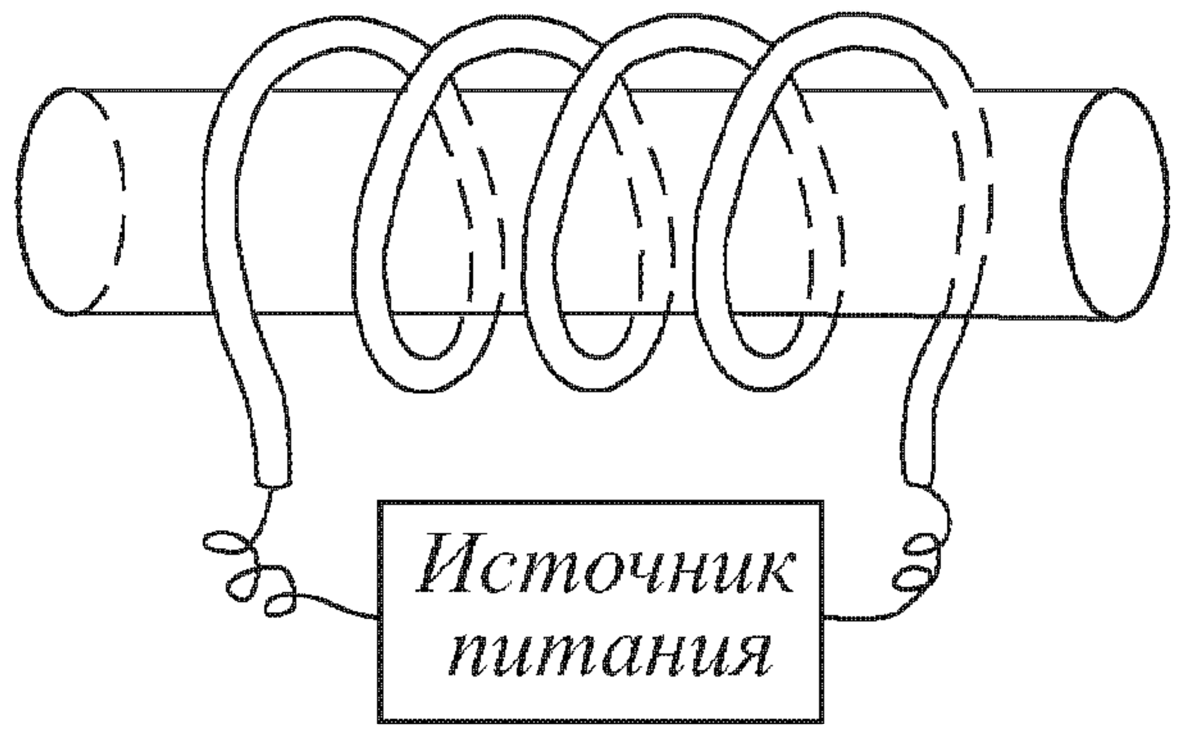


Рис. 347

Если бы энергетический спектр состоял только из двух уровней $\mathcal{E}_1$ и $\mathcal{E}_2$, то с помощью оптической накачки создать инверсную заселенность их было бы невозможно. Действительно, как видно из (119.2), индуцированные переходы атомов с нижнего уровня на верхний и обратно идут с одинаковыми скоростями. Поэтому лампа-вспышка самое большее могла бы лишь уравнять заселенности обоих уровней. Наличие же спонтанного излучения приводило бы к обеднению (атомами) верхнего уровня по сравнению с нижним.

Положение меняется благодаря наличию третьего, широкого энергетического уровня, состоящего из полос $\mathcal{E}_3$ и $\mathcal{E}_4$. Лампа-вспышка переводит атомы хрома из невозбужденного состояния в возбужденное, т.е. в энергетические полосы $\mathcal{E}_3$ и $\mathcal{E}_4$. Значительная ширина этих полос имеет большое практическое значение.

Лампа-вспышка излучает свет, близкий к белому. Если бы уровни $\mathcal{E}_3$ и $\mathcal{E}_4$ были очень узкими, то лишь ничтожная часть энергии лампы-вспышки могла бы быть использована на их возбуждение. Благодаря же значительной ширине полос $\mathcal{E}_3$ и $\mathcal{E}_4$ на их возбуждение уходит 10–15 % лучистой энергии лампы-вспышки. На уровнях $\mathcal{E}_3$ и $\mathcal{E}_4$ возбужденные атомы хрома живут очень короткое время ($\sim 10^{-8}$ с). За это время они переходят на один из уровней $\mathcal{E}_2$. При таком переходе атомы хрома не излучают, а расходуют свою энергию на возбуждение колебаний кристаллической решетки. Возможность обратного возвращения атома из полос $\mathcal{E}_3$ и $\mathcal{E}_4$ на уровень $\mathcal{E}_1$, хотя и существует, но вероятность такого процесса пренебрежимо мала по сравнению с вероятностью перехода на уровни $\mathcal{E}_2$. Уровни $\mathcal{E}_2$ метастабильны. Время жизни атома на них $\sim 10^{-3}$ с, что по атомным масштабам очень велико. Это позволяет накапливать атомы на уровнях $\mathcal{E}_2$. Если переводить атомы с уровня $\mathcal{E}_1$ на уровни $\mathcal{E}_3$ и $\mathcal{E}_4$ достаточно быстро, то на уровни $\mathcal{E}_2$ можно перевести больше половины атомов хрома. Тогда на уровнях $\mathcal{E}_2$ окажется больше атомов, чем на уровне $\mathcal{E}_1$, т.е. возникнет инверсная заселенность этих уровней.

3. Подсчитаем мощность лампы-вспышки, необходимую для создания инверсной заселенности уровней $\mathcal{E}_1$ и $\mathcal{E}_2$. Средняя длина волны, излучаемая при переходе из полос $\mathcal{E}_3$ и $\mathcal{E}_4$ на уровень $\mathcal{E}_1$, порядка 450 нм. Поэтому для перевода одного атома хрома с уровня $\mathcal{E}_1$ на уровни $\mathcal{E}_3$ и $\mathcal{E}_4$ требуется в среднем энергия $h\nu = hc/\lambda \approx 4{,}4 \cdot 10^{-12}$ эрг. В рубине, обычно употребляемом для лазеров, на каждый см^3 приходится около 10^{19} атомов хрома. Не менее половины из них надо перевести в полосы $\mathcal{E}_3$ и $\mathcal{E}_4$. На это требуется энергия $\approx 2{,}2 \cdot 10^7$ эрг $= 2{,}2$ Дж. Атомы должны быть переведены на возбужденные уровни $\mathcal{E}_3$ и $\mathcal{E}_4$ за время, не превышающее времени жизни атома хрома на уровне $\mathcal{E}_2$ ($\sim 2 \cdot 10^{-3}$ с). В противном случае инверсная заселенность получиться не может. Таким образом, минимальная мощность, идущая на возбуждение атомов, должна быть порядка $2{,}2/10^{-3} = 2200$ Вт/см^3 $= 2{,}2$ кВт/см^3. Если коэффициент использования лучистой энергии лампы принять равным 10–15 %, то потребляемая мощность лампы должна быть ~ 20 кВт/см^3. Для рубинового стержня с объемом 10 см^3 получится ~ 200 кВт. Это достигается сравнительно легко. Нетрудно подсчитать, что при напряжении 2000 В емкость батареи конденсаторов должна быть не меньше 100 мкФ. Обычно употребляются батареи с емкостью ~ 1000 мкФ.

Значительная доля лучистой энергии (более 50 %), поглощенной рубиновым стержнем, тратится на его нагревание. Но при температурах, превышающих примерно 1000 К, рубиновый лазер перестает генерировать. Поэтому в некоторых конструкциях лазеров предусмотрено охлаждение рубина проточной водой и даже жидким азотом.

4. Рубин — *одноосный кристалл.* Обыкновенный и необыкновенный показатели преломления светло-красного рубина, употребляемого в лазерах, равны соответственно $n_o = 1{,}7653$ и $n_e = 1{,}7573$ (для $\lambda = 656$ нм). Рубиновый лазер может давать поляризованный свет без каких-либо специальных поляризационных приспособлений. Для этого оптическая ось рубинового стержня не должна совпадать с его геометрической осью. Возникновение линейной поляризации нельзя объяснить различием коэффициентов отражения обыкновенного и необыкновенного лучей, так как разность $n_o - n_e$ слишком мала. Поляризация возникает потому, что в обыкновенной волне направления волновой нормали и луча совпадают между собой, а в необыкновенной не совпадают. Чтобы зеркала лазера действовали эффективно, как резонатор, необходимо, чтобы волновые нормали были к ним перпендикулярны. Но тогда в рубине только обыкновенный луч будет распространяться параллельно геометрической оси стержня, а необыкновенный пойдет под углом к ней, попадая на боковую поверхность стержня. Поэтому резонатор будет эффективнее усиливать обыкновенные лучи, чем необыкновенные. Из лазера будут преимущественно выходить обыкновенные лучи, в которых электрический вектор перпендикулярен к оптической и геометрической осям кристалла. Если эти оси совпадают, то поляризация не возникает. Если же, как показали наблюдения, угол между ними заключен в пределах от 60 до 90°, то линейная поляризация полная.

5. Наиболее распространенным является *импульсный режим* работы рубинового лазера. Лампа-вспышка дает импульс света длительностью $\sim 10^{-3}$ с. Длительность когерентного импульса, излучаемого лазером, несколько короче. Это связано, во-первых, с тем, что требуется некоторое время, чтобы заселенность уровня $\mathcal{E}_2$ достигла порогового значения, после чего лазер начинает генерировать. Во-вторых, с тем, что генерация лазера прекращается, когда интенсивность вспышки лампы понижается настолько, что ее становится уже недостаточно для поддержания заселенности выше пороговой. Сам лазерный импульс имеет сложную структуру и состоит из множества отдельных импульсов длительностью до 10^{-6} с, следующих друг за другом с интервалом $\sim 3 \cdot 10^{-6}$–10^{-5} с. Мощность рубинового лазера в импульсе может достигать десятков кВт при длине стержня в 20–30 см и диаметре 1,5 см.

Эта мощность может быть повышена до нескольких десятков мегаватт за счет уменьшения длительности генерируемого импульса. Для этого на уровень $\mathcal{E}_2$ надо перевести не половину, а большую часть атомов хрома, оставив уровень $\mathcal{E}_1$ практически незаполненным. Тогда при длине рубинового стержня 10–20 см импульс генерируемого света, как показывает расчет, может развиться за время $\sim 10^{-8}$ с. Лампа-вспышка легко могла бы произвести такой переход, если бы в рубине не возникала генерация, возвращающая атомы хрома снова на исходный уровень $\mathcal{E}_1$. А генерация начинается сразу же, как только заселенность уровня $\mathcal{E}_2$ станет выше пороговой. Задачу можно решить, если на короткое время выключить обратную связь, осуществляемую зеркалами. За это время можно перевести все атомы хрома на возбужденный уровень $\mathcal{E}_2$. При выключенных зеркалах среднее время жизни атома хрома на уровне $\mathcal{E}_2$ составляет около 0,002 с. Если в течение этого времени на мгновение снова включить зеркала, то происходит почти мгновенный вынужденный переход атомов на невозбужденный уровень $\mathcal{E}_1$ и связанная с ним мощная вспышка лазерного излучения.

6. Быстрое включение и выключение обратной связи, а также вообще изменение добротности резонатора можно осуществить, заменив одно из зеркал призмой P полного отражения (рис. 348), вращающейся с угловой скоростью ~ 500 об/с. Лазер может генерировать только в положении призмы, изображенном на рис. 348, или очень близком к нему. Во всех остальных положениях обратной связи нет и лазер не генерирует. В это время происходит оптическая накачка рубинового стержня R. Когда призма займет положение, указанное на рис. 348, происходит излучение мощного лазерного импульса. Хотя при этом призма и продолжает быстро вращаться, но за короткое время $\sim 10^{-8}$ с, требующееся для развития лазерного импульса, она не успевает сколько-нибудь заметно сместиться из указанного положения. После высвечивания запасенной энергии при дальнейшем вращении призмы снова начинается оптическая накачка, пока при повороте на 360° не произойдет излучение нового лазерного импульса. Однако при прохождении через стержень мощного импульса выключение призмы не всегда достаточно для прекращения действия обратной связи, так как она может сохраняться из-за отражения света от концов рубинового стержня (для рубина коэффициент отражения $(n-1)^2/(n+1)^2 \approx 7{,}6\,\%$). Во избежание этого концы рубинового

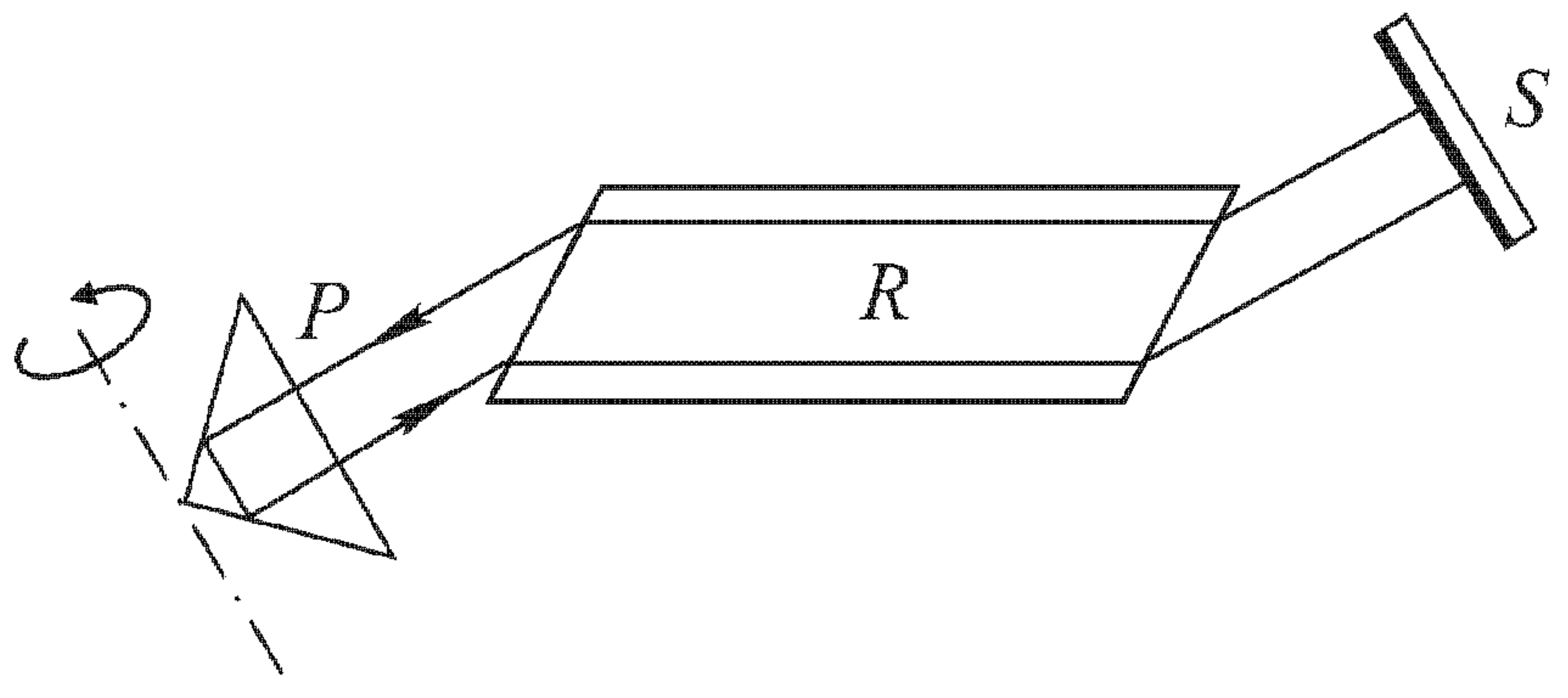


Рис. 348

стержня делают скошенными под углом к его геометрической оси. Обычно угол падения берут равным *углу Брюстера* (на границе рубин–воздух $\varphi_B \approx 30^\circ$). Тогда генерируется линейно поляризованный свет, электрический вектор которого лежит в плоскости падения.

7. Более быстрое включение и выключение обратной связи производится с помощью оптического затвора — *ячейки Керра* с нитробензолом или *ячейки Поккельса* с анизотропным кристаллом (рис. 349). Ячейка вместе со своим конденсатором повернута вокруг оси установки на угол 45° (это на рис. 349 не показано). Поляризатор P (поляроид или поляризационная призма) ориентирован так, чтобы полностью пропускать линейно поляризованное излучение рубинового стержня, когда он генерирует. Перед включением лампы-вспышки на ячейку K подается такое напряжение, чтобы она поворачивала плоскость поляризации излучения лазера на угол 90°. Тогда свет через поляризатор P не пройдет, т.е. обратная связь будет выключена. В то время, когда лампа дает вспышку, происходит оптическая накачка рубинового стержня. Если затем быстро снять напряжение с конденсатора ячейки, то линейно поляризованное излучение рубина начнет свободно распространяться между зеркалами S_1 и S_2 и возникает мощный импульс индуцированного излучения лазера длительностью порядка 10^{-8} с.

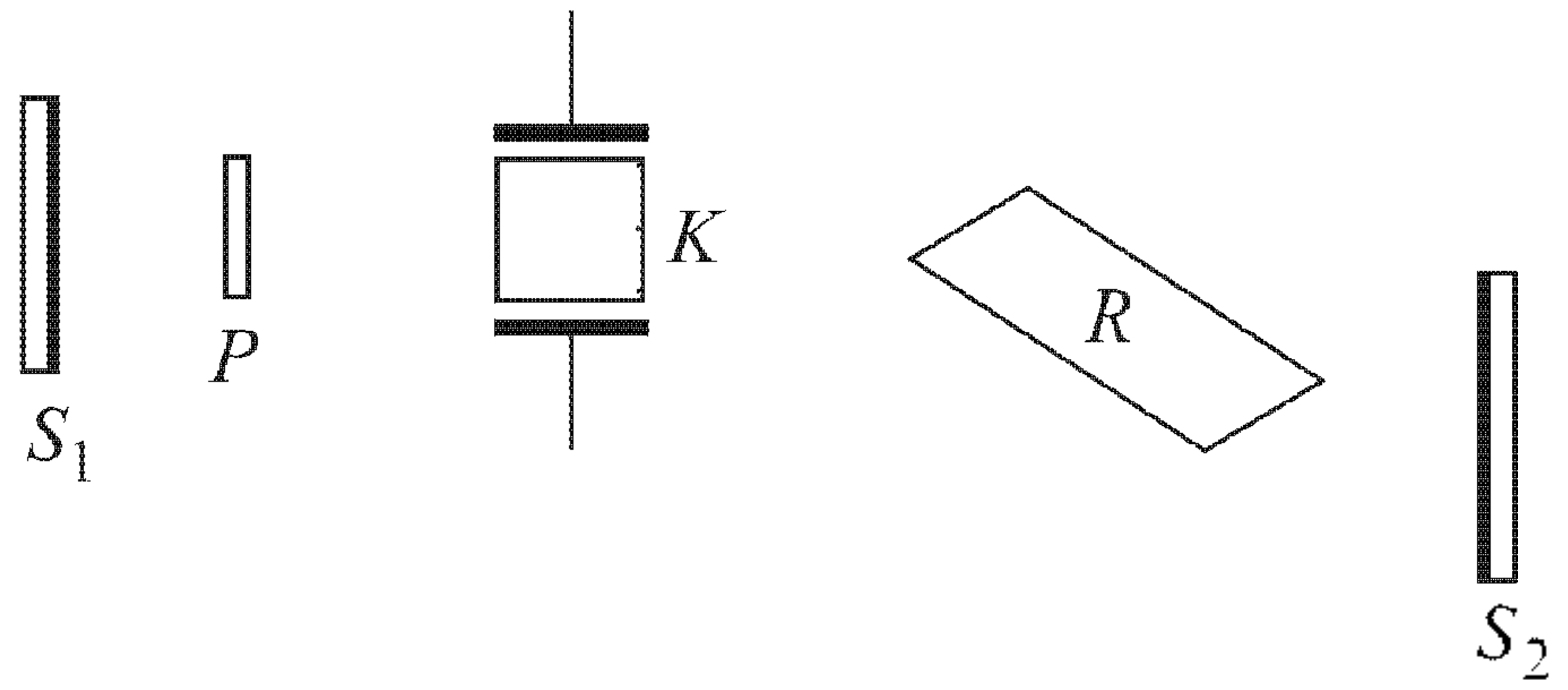


Рис. 349

8. Большое распространение в качестве оптических затворов получили также *просветляющиеся фильтры*, помещаемые вместо ячейки Керра или Поккельса. (Поляризатор P в этом случае не нужен.) Их действие основано на увеличении прозрачности вещества, когда интенсивность света становится достаточно большой, как это имеет место в случае излучения лазеров (см. § 89, п. 6). При малой интенсивности света фильтр поглощает свет, почти полностью устраняя обратную связь. С увеличением заселенности верхнего уровня возникает слабая генерация рубинового стержня, несколько уменьшающая поглощение фильтра. Это приводит к усилению обратной связи и вызывает лавину лазерного излучения. Последняя по мере нарастания все более и более просветляет фильтр. Когда интенсивность излучения начнет уменьшаться, поглощение фильтра будет быстро возрастать, а обратная связь ослабляться. Поскольку вся система работает автоматически, лазер с просветляющимися фильтрами во время вспышки лампы накачки может генерировать серию импульсов, следующих друг за другом.

9. Быстрое включение обратной связи позволяет сократить длительность импульса лазерного излучения до 10^{-7}–10^{-8} с. Так как энергия импульса, снимаемая с рубинового стержня длиной 20 см и диаметром 1,5 см, составляет 1–2 Дж, то при этом развивается мощность 10–200 МВт. Дальнейшее повышение мощности лазерного импульса может быть достигнуто путем усиления света в каскаде из нескольких последовательно расположенных лазеров. Первый лазер в таком каскаде должен служить *генератором*, остальные — *усилителями света*. Если мощность генерируемого импульса достаточно велика, то уже небольшой его части вблизи передового фронта достаточно, чтобы вызвать вынужденные переходы в невозбужденное состояние всех возбужденных атомов хрома в рубине усилителей. Это сокращает длительность импульса и повышает его мощность. Таким путем удалось получить *гигантские импульсы света* длительностью в $2 \cdot 10^{-9}$ с при полной энергии импульса $\sim$ 20 Дж. Это соответствует мощности $\sim 10^4$ МВт. Но и такая мощность еще не предел. Лазер с просветляющимся фильтром дает импульс длительностью $\sim 10^{-8}$ с, состоящий из последовательности импульсов, длительность которых может составлять 10^{-1}–10^{-12} с. Выделение одного такого импульса с последующим усилением его позволяет достигнуть мощности $\sim 10^7$ МВт.

10. Из других твердых материалов, употребляемых в лазерах, надо упомянуть прежде всего *неодимовое стекло* и *флюорит кальция* (CaF_2) с небольшой примесью атомов редкоземельных элементов: *диспрозия* (Dy), *самария* (Sm) и пр. Активность среды создается ионами неодима Nd^{+++}, диспрозия Dy^{++}, самария Sm^{++} и т.д. Неодимовое стекло генерирует инфракрасное излучение с длиной волны $\lambda = 1{,}06$ мкм, а флюорит кальция с диспрозием — еще более длинноволновое излучение с $\lambda = 2{,}36$ мкм. Стержни из неодимового стекла изготовить легче, чем из рубина. Стеклянные стержни могут достигать длины 0,5–1 м при диаметре до 5 см. Ширина полос у неодимового стекла больше, чем у рубина, что позволяет получить больший КПД. Зато у рубина более высокая прочность. При длительности импульса 10^{-8}–10^{-9} с рубин выдерживает мощности $(3\text{–}4)\cdot 10^3$ МВт на каждый см^2 поперечного сечения стержня, а неодимовое стекло — только $(1\text{–}1{,}5)\cdot 10^3$ МВт.

В отличие от рубинового лазера, работающего по *трехуровневой схеме*, лазер на флюорите кальция с примесью диспрозия работает по *четырехуровневой схеме*. Это позволяет создавать достаточно мощные лазеры, работающие *не импульсами, а* *непрерывно*, что для некоторых применений очень существенно. Между основным уровнем $\mathcal{E}_1$ диспрозия в кристалле и энергетической полосой $\mathcal{E}_2$ расположены два промежуточных уровня $\mathcal{E}_3$ и $\mathcal{E}_4$ (рис. 350). Между ними и создается инверсная заселенность атомов диспрозия. Лампа-вспышка переводит атомы диспрозия через широкую энергетическую полосу $\mathcal{E}_2$ на уровень $\mathcal{E}_3$. Уровень $\mathcal{E}_4$ по энергии приподнят над уровнем $\mathcal{E}_1$ настолько высоко, что при температуре жидкого азота его заселенность N_4 практически равна нулю. Поэтому нет необходимости переводить на уровень $\mathcal{E}_3$ больше половины атомов диспрозия. Достаточно лишь, чтобы разность заселенностей $N_3 - N_4$ уровней $\mathcal{E}_3$ и $\mathcal{E}_4$ превысила пороговое значение. Оно составляет лишь небольшую долю атомов диспрозия. Достаточна сравнительно небольшая мощность лампы-накачки ($\approx$ 15–20 Вт), чтобы осуществить непрерывную работу лазера. Кроме того, непрерывная накачка позволяет генерировать гигантские импульсы с большой частотой повторения (до нескольких кГц). Мощности света, получаемые при этом в каждом импульсе, около 1–2 МВт.

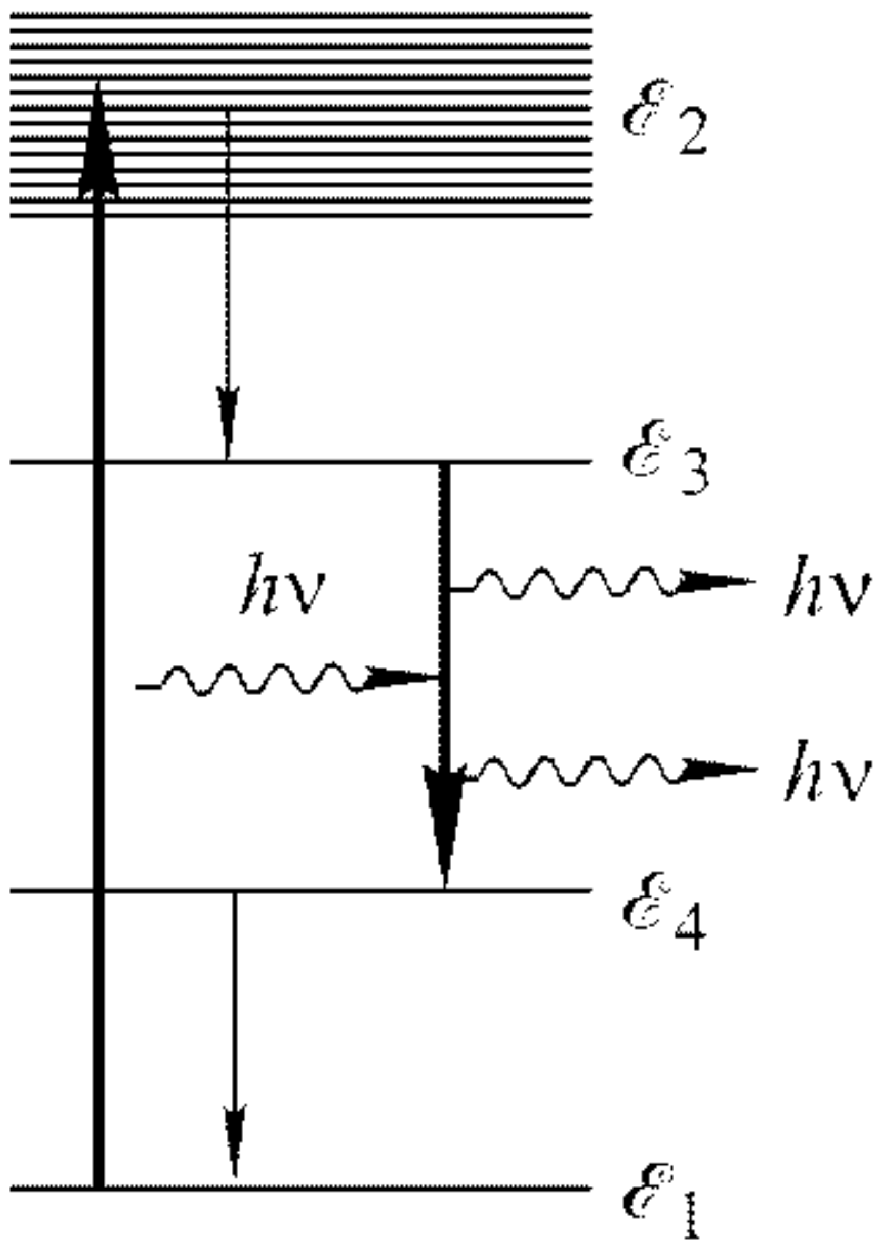


Рис. 350

ЗАДАЧИ

1. Мощность рубинового лазера в импульсе равна $W = 100$ МВт. Площадь поперечного сечения рубинового стержня $S = 1$ см^2. Оценить напряженность E электрического поля и давление света в лазерном пучке, считая его строго плоскопараллельным ($\lambda = 694{,}3$ нм). Как изменятся эти величины, если пучок сфокусировать идеальной линзой с фокусным расстоянием $f = 5$ см?

Решение. Напряженность электрического поля E_0 в несфокусированном пучке оценим по формуле

$$W = \frac{c}{4\pi}\overline{EH}S = \frac{c}{4\pi}\overline{E^2}S,$$

а давление излучения $\mathcal{P}_0$ — по формуле $\mathcal{P}_0 = W/(cS)$. Таким путем находим

$$E_0 \approx \sqrt{\overline{E^2}} = \sqrt{\frac{4\pi W}{cS}} = 647 \text{ СГСЭ} = 1{,}94 \cdot 10^5 \text{ В/см},$$

$$\mathcal{P}_0 \approx 3{,}3 \cdot 10^4 \text{ дин/см}^2 \approx 0{,}033 \text{ атм}.$$

Для расчета соответствующих величин в фокусе можно пользоваться теми же формулами, вычислив предварительно ширину пучка. Для оценки будем считать, что весь свет концентрируется в пределах центрального светлого кружка с радиусом $R = 0{,}61 f\lambda/r$ и площадью

$$\pi R^2 = \pi(0{,}61 f\lambda/r)^2 = (0{,}61\pi f\lambda)^2/S,$$

где r — радиус поперечного сечения падающего пучка. Эту площадь и надо подставить в предыдущие формулы вместо S. В результате получим

$$E \approx \sqrt{\overline{E^2}} \approx \frac{S}{0{,}61\pi f\lambda} E_0 = 1{,}5 \cdot 10^3 E_0 = 2{,}9 \cdot 10^8 \text{ В/см},$$

$$\mathcal{P} = \left(\frac{S}{0{,}61\pi f\lambda}\right)^2 \mathcal{P}_0 = 2{,}25 \cdot 10^6 \mathcal{P}_0 \approx 7{,}4 \cdot 10^4 \text{ атм}.$$

Такое давление было бы, если бы пучок лазерного света фокусировался в вакууме. Если же он падает на вещество, то из-за большой концентрации энергии вблизи фокуса происходит практически мгновенное испарение вещества (с превращением его в плазму). С этим связано еще большее повышение давления.

2. Эффективной температурой $T_{\text{эф}}$ лазерного излучения называется такая температура абсолютно черного тела, при которой оно дает излучение той же удельной интенсивности I_ν частоты ν, что и лазер. Оценить эффективную температуру гелий-неонового лазера, генерирующего в непрерывном режиме свет с длиной волны $\lambda = 632{,}8$ нм ($\nu \approx 5 \cdot 10^{14}$ Гц.) Ширина спектральной линии генерируемого света $\delta\nu \approx 10^4$ Гц, расходимость светового пучка $\delta\vartheta \approx 1' \approx 3 \cdot 10^{-4}$ рад, мощность излучения $W = 10$ мВт.

Решение. Удельная интенсивность лазерного излучения равна $I_\nu = W/(\delta\nu \cdot \delta\Omega)$, где $\delta\Omega = (\delta\vartheta)^2 \approx 10^{-7}$ ср — телесный угол, определяющий расходимость светового пучка. По формуле Рэлея–Джинса

$$I_\nu = \frac{2k\nu^2}{c^2} T_{\text{эф}}.$$

Приравнивая это предыдущему выражению, получим

$$T_{\text{эф}} \approx \frac{c^2 W}{2k\nu^2 \delta\nu \cdot \delta\Omega} \approx 10^{15} \text{ К},$$

что примерно в 10^{11} раз превосходит температуру Солнца.

§ 122. Гелий-неоновый лазер

1. Энергетический спектр газов, во всяком случае при тех давлениях, которые применяются в рекламных трубках и лазерах, весьма точно совпадает с энергетическим спектром *изолированных атомов и молекул*, из которых состоит газ. Поэтому спектральные линии газов *более узкие*, чем линии примесей в твердых телах. Газы отличаются *высокой оптической однородностью* и *малой плотностью*, а поэтому слабо рассеивают и искажают волны, распространяющиеся в них. Все это позволяет использовать в газовых лазерах большие расстояния между зеркалами и получать *острую направленность*, *высокую монохроматичность* и *стабильность частоты излучения*. Так, в гелий-неоновом лазере реально получаемая расходимость пучка лучей обычно составляет 1–2$'$, а в лучших случаях достигает теоретического предела $\delta\vartheta \sim \lambda/D \sim 20''$. Относительная ширина спектральной линии в излучении такого лазера будет равна $\delta\omega/\omega \sim 10^{-11}$–$10^{-12}$, а сама частота ω в оптимальных условиях поддерживается с относительной точностью $\sim 10^{-14}$ (хотя в обычных условиях относительная точность не выше 10^{-10}).

Однако при малой плотности газов в них нельзя получить столь же большие концентрации возбужденных атомов, а потому и столь же большие импульсные мощности излучения, как в твердых телах. Так, выходная мощность гелий-неонового лазера в непрерывном режиме обычно составляет от десяти до нескольких сот милливат. Однако, ввиду высокой монохроматичности и направленности излучения, эта величина все же громадна по сравнению с тем, что могут дать тепловые источники света. Она соответствует эффективной температуре излучения, превышающей температуру Солнца примерно в 10^{11}–10^{12} раз (см. задачу 2 к предыдущему параграфу). Впрочем, в непрерывном режиме инфракрасный лазер на CO_2 может генерировать до 10 кВт, а ионный аргоновый лазер в видимой области — до ~ 1 кВт. В импульсном режиме мощность этих лазеров может составлять несколько сот киловатт.

2. Первым газовым лазером был *гелий-неоновый лазер*, созданный в конце 1960 г. Джаваном (р. 1926), Беннетом (р. 1903) и Эрриотом. Принципиальная схема гелий-неонового лазера в ее современном виде приведена на рис. 351. Лазер состоит из газоразрядной трубки T длиной от нескольких десятков сантиметров до 1,5–2 м и внутренним диаметром 7–10 мм. Трубка наполнена смесью гелия (давление ~ 1 мм рт. ст.) и неона ($\sim 0{,}1$ мм рт. ст.). Концы трубки закрыты плоскопараллельными стеклянными или кварцевыми пластинками P_1 и P_2, установленными под углом Брюстера к ее оси. Это создает линейную поляризацию лазерного излучения с электрическим вектором, параллельным плоскости падения. Зеркала S_1 и S_2, между которыми помещается трубка, делаются обычно сферическими с многослойными диэлектрическими покрытиями. Они имеют высокие коэффициенты отражения и практически не поглощают свет. Пропускаемость зеркала, через которое преимущественно

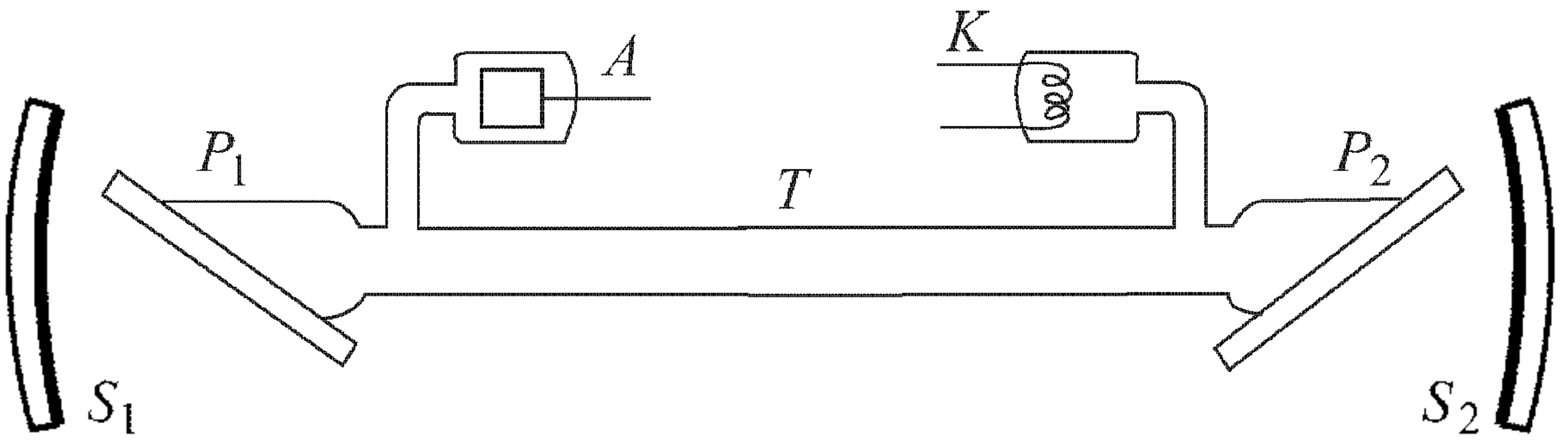


Рис. 351

выходит излучение лазера, составляет обычно 2 %, другого — менее 1 %. Между электродами трубки прикладывается постоянное напряжение 1–2 кВ. Катод K трубки может быть холодным, но для увеличения разрядного тока применяют также трубки с пустотелым цилиндрическим анодом, катод которых нагревается низковольтным источником тока. Разрядный ток в трубке составляет несколько десятков миллиампер. В первом варианте гелий-неонового лазера, построенном Мейманом, применялась безэлектродная трубка, в которой возбуждался высокочастотный разряд. Лазер может работать и в непрерывном, и в импульсном режиме. Он генерирует красный свет с длиной волны 632,8 нм и может генерировать также инфракрасное излучение с длинами волн 1,150 и 3,390 мкм. Но тогда необходимо иметь торцевые окна, прозрачные для инфракрасного света, и зеркала с высокими коэффициентами отражения в инфракрасной области спектра.

3. Рассмотрим теперь, как возникает инверсная заселенность атомов неона. Упрощенная схема уровней неона приведена справа на рис. 352. Выше уровня $\mathcal{E}_4$ у неона имеется еще 28 уровней с энергией, меньшей $\mathcal{E}_3$, но они для нас не имеют значения и на рисунке не изображены. Возбуждение атомов неона происходит в результате *столкновений их с электронами газоразрядной плазмы.* При определенном режиме разряда этот процесс может привести к инверсной заселенности уровней $\mathcal{E}_1$ и $\mathcal{E}_2$. Однако заселенность уровней $\mathcal{E}_1$ и $\mathcal{E}_3$, а также уровней $\mathcal{E}_4$ и $\mathcal{E}_3$ остается неинверсной. Инверсной заселенности препятствует долгоживущий метастабильный уровень $\mathcal{E}_5$, лежащий немного ниже короткоживущего уровня $\mathcal{E}_1$. Заселенность уровня $\mathcal{E}_5$ велика, за счет этого происходит пополнение быстро опустошающегося уровня $\mathcal{E}_1$, и инверсии заселенности между уровнями $\mathcal{E}_1$ и $\mathcal{E}_3$ не возникает.

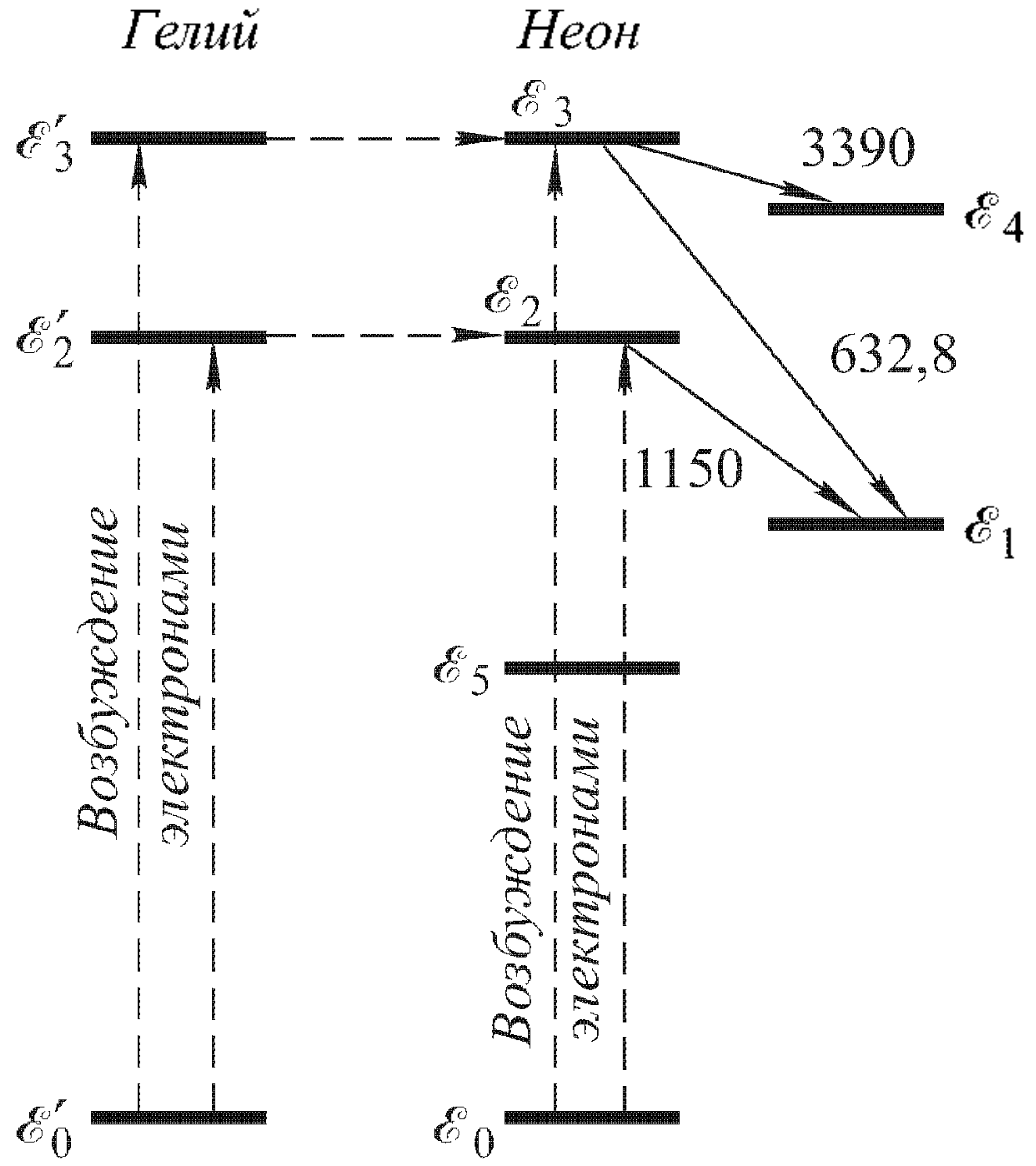


Рис. 352

Добавление гелия меняет дело. Из всех уровней гелия, помимо нормального $\mathcal{E}'_0$, для работы лазера имеют значение метастабильные уровни $\mathcal{E}'_2$ и $\mathcal{E}'_3$ с энергиями 19,82 и 20,61 эВ соответственно. Спонтанный радиационный переход с этих уровней на основной уровень $\mathcal{E}'_0$ «запрещен», т.е. происходит с очень малой вероятностью. Поэтому время жизни атома на уровнях $\mathcal{E}'_2$ и $\mathcal{E}'_3$ очень велико. В результате электронных ударов на этих метастабильных уровнях накапливается очень много атомов гелия. Но уровни гелия $\mathcal{E}'_2$ и $\mathcal{E}'_3$ почти совпадают с уровнями $\mathcal{E}_2$ и $\mathcal{E}_3$ неона. Благодаря этому при столкновениях возбужденных атомов гелия с невозбужденными атомами неона интенсивно происходят *безызлучательные переходы* атомов гелия в невозбужденное состояние с резонансной передачей энергии атомам неона. Этот процесс возбуждения атомов неона на рис. 352 символически изображен горизонтальными штриховыми стрелками. В результате концентрации атомов неона на уровнях $\mathcal{E}_2$ и $\mathcal{E}_3$ сильно возрастают, и возникает инверсная заселенность по отношению к уровням $\mathcal{E}_1$ и $\mathcal{E}_3$, а разность заселенностей уровней $\mathcal{E}_2$ и $\mathcal{E}_1$ увеличивается в несколько раз.

Выясним в заключение влияние *столкновений* атомов неона со стенками трубки. Такие столкновения практически не влияют на заселенность уровней $\mathcal{E}_2$ и $\mathcal{E}_3$ и непосредственно уровня $\mathcal{E}_1$, так как все эти уровни *короткоживущие.* За время жизни в возбужденных состояниях на этих уровнях атомы неона практически не успевают доходить до стенок трубки. Указанные уровни разрушаются значительно раньше. Напротив, на уровне $\mathcal{E}_5$ возбужденные атомы живут долго, претерпевая в этих состояниях многочисленные столкновения со стенками трубки. Столкновения разгружают уровень $\mathcal{E}_5$, в результате чего атомы неона переходят с уровня $\mathcal{E}_1$ на уровень $\mathcal{E}_5$. Опустошение уровня $\mathcal{E}_1$ происходит быстрее, чем при заселенном уровне $\mathcal{E}_5$. Разница заселенностей уровней $\mathcal{E}_3$ и $\mathcal{E}_1$ увеличивается, что повышает эффективность работы лазера. Процесс опустошения уровня $\mathcal{E}_1$ происходит наиболее эффективно при некотором *оптимальном диаметре трубки.* Опыты показали, что максимальная мощность гелий-неонового лазера достигается при диаметре трубки ~ 7 мм. При бо́льших диаметрах мощность лазера падает, несмотря на сильное увеличение объема рабочего газа (объем трубки пропорционален квадрату ее диаметра). Это связано с тем, что эффективное опустошение уровня $\mathcal{E}_1$ происходит у атомов, находящихся *вблизи стенок трубки*, а атомы, находящиеся вблизи ее центра, *практически выключаются из процесса генерации.*

§ 123. Нелинейная поляризация среды

1. Изобретение лазеров сделало возможным экспериментировать с интенсивными световыми пучками, в которых напряженность электрического поля не пренебрежимо мала по срав-

можно пользоваться моделью изотропной среды, полагая

$$\mathbf{P} = \alpha\mathbf{E} + \alpha_2 E\mathbf{E} + \alpha_3 E^2\mathbf{E} + \ldots, \qquad (123.2)$$

где поляризуемости α, α_2, α_3, ... являются уже *скалярами.* Такое упрощение вполне допустимо *при качественном рассмотрении* возможных нелинейных оптических явлений. Надо только иметь в виду, что в кристаллах в выбранном направлении могут распространяться волны не всех, а только *избранных поляризаций.* Соотношение (123.2) приближенно применимо к каждой из таких волн, причем для различных волн поляризуемости α, α_2, ... имеют разные значения. Кроме того, волны разных поляризаций могут *нелинейно взаимодействовать*, обмениваясь энергией друг с другом. Такое взаимодействие должно иметь место при тензорной связи (123.1) между **P** и **E**. Но оно было бы невозможно, если бы эта связь была скалярной типа (123.2). Понятно, что при нашем подходе влияние такого взаимодействия может быть учтено только качественно.

2. Разобьем поляризацию **P** на *линейную* и *нелинейную части*: $\mathbf{P} = \mathbf{P}_{\text{л}} + \mathbf{P}_{\text{нл}}$. Нелинейная часть определяется выражением

$$\mathbf{P}_{\text{нл}} = \alpha_2 E\mathbf{E} + \alpha_3 E^2\mathbf{E} + \ldots, \qquad (123.3)$$

а линейная $\mathbf{P}_{\text{л}} = \alpha\mathbf{E}$. В соответствии с этим и индукция $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$ представится суммой *линейной части* $\mathbf{D}_{\text{л}} = \mathbf{E} + 4\pi\mathbf{P}_{\text{л}}$ и *нелинейной* $\mathbf{D}_{\text{нл}} = 4\pi\mathbf{P}_{\text{нл}}$. Линейная часть, очевидно, равна $\mathbf{D}_{\text{л}} = \varepsilon\mathbf{E}$, где ε — обычная *диэлектрическая проницаемость среды*, как она определяется в линейной электродинамике. После этого запишем систему фундаментальных уравнений Максвелла в следующем виде:

$$\begin{gathered}
\operatorname{rot}\mathbf{H} - \frac{\varepsilon}{c}\frac{\partial\mathbf{E}}{\partial t} = \frac{4\pi}{c}\frac{\partial\mathbf{P}_{\text{нл}}}{\partial t},\\
\operatorname{rot}\mathbf{E} - \frac{1}{c}\frac{\partial\mathbf{H}}{\partial t} = 0,\\
\operatorname{div}(\varepsilon\mathbf{E}) = -4\pi\operatorname{div}\mathbf{P}_{\text{нл}},\\
\operatorname{div}\mathbf{H} = 0.
\end{gathered} \qquad (123.4)$$

Для решения такой системы применяем *метод последовательных приближений.* В нулевом приближении в уравнении (123.4) отбрасываем правые части. Получатся *обычные уравнения линейной электродинамики.* В качестве нулевого приближения возьмем плоскую волну

$$\mathbf{E} = \mathbf{E}_0 = \mathbf{A}\cos(\omega t - \mathbf{kr}), \qquad (123.5)$$

где волновой вектор **k** удовлетворяет обычному соотношению $\mathbf{k}^2 = \varepsilon\omega^2/c^2$. Для нахождения первого приближения в (123.3) отбросим кубичные и высшие члены, а в квадратичном члене

$\alpha_2 E\mathbf{E}$ поле $\mathbf{E}$ заменим его выражением (123.5) в нулевом приближении. После этого снова получатся *линейные уравнения*, но уже *неоднородные*, с известными правыми частями. Эти правые части могут быть истолкованы как *добавочные источники волн*, обусловленные нелинейной частью поляризации среды. Каждый элемент объема dV среды *переизлучает волны* как диполь Герца с добавочным дипольным моментом $\mathbf{P}_{\text{нл}}\,dV$. Эти излучения, накладываясь на волну (123.5), и создают волновое поле $\mathbf{E}_0+\mathbf{E}_1$ *в первом приближении*. Второе приближение находится так же. Для этого выражение (123.3) обрываем на членах третьей степени, заменяя в оборванном выражении вектор $\mathbf{E}$ на $\mathbf{E}_0+\mathbf{E}_1$, после чего вычисляем поле $\mathbf{E}_0+\mathbf{E}_1+\mathbf{E}_2$ *во втором приближении*, и т.д.

К изложенному надо еще добавить, что следует понимать под ε в уравнениях (123.4), когда среда *обладает дисперсией*. Ответ заключается в следующем. Если взять какое-либо приближение, то в правой части уравнений (123.4) появятся слагаемые не только с исходной частотой ω, но и с частотами 2ω, $3\omega, \ldots$ Могут появиться и другие частоты, как, например, при параметрической генерации света (см. § 126). Надо $\mathbf{E}$ и $\mathbf{H}$ искать в виде суммы монохроматических полей *с теми же частотами*. Уравнение (123.4) следует написать *для каждой частоты в отдельности*, сохранив в правой части только члены той же частоты и понимая под ε значение функции ε также *при той же частоте*. Именно такой *символический смысл* имеет система уравнений (123.4).

§ 124. Первое приближение. Оптическое детектирование. Генерация вторых гармоник, суммарной и разностной частот

1. Нелинейная добавка (123.3) к поляризации среды, вычисленная в нулевом приближении, равна

$$P_{\text{нл}} = \alpha_2 E_0^2 = \frac{\alpha_2 A^2}{2} + \frac{\alpha_2 A^2}{2}\cos\left[2(\omega t - \mathbf{kr})\right]. \qquad (124.1)$$

Первое слагаемое в этом выражении не зависит от времени. С ним связано так называемое *оптическое детектирование*, т.е. возникновение в нелинейной среде постоянной электрической поляризации при прохождении через нее мощной световой волны. Это явление аналогично *выпрямлению синусоидального электрического тока*. Его можно наблюдать, если между обкладками конденсатора, одна из которых заземлена через большое сопротивление, поместить кристалл (например, кварца) и пропустить через него световой пучок от рубинового лазера.

Вследствие детектирования световой пучок возбуждает в цепи конденсатора *импульс электрического тока*, который можно обнаружить с помощью осциллографа.

2. Второе слагаемое в (124.1) гармонически меняется во времени. С ним связана *генерация в нелинейной среде второй гармоники*, т.е. волны с удвоенной частотой $\omega_2 = 2\omega$. Для нахождения поля этой гармоники поступаем так, как изложено в предыдущем параграфе. Переходя к комплексной форме, получаем систему уравнений

$$\begin{aligned} \operatorname{rot}\mathbf{H} - \frac{\varepsilon(2\omega)}{c}\frac{\partial\mathbf{E}}{\partial t} &= i\omega\frac{4\pi\alpha_2}{c}A\mathbf{A}e^{2i(\omega t-\mathbf{kr})}, \\ \operatorname{rot}\mathbf{E} - \frac{1}{c}\frac{\partial\mathbf{H}}{\partial t} &= 0, \\ \operatorname{div}\mathbf{E} = \operatorname{div}\mathbf{H} &= 0. \end{aligned} \tag{124.2}$$

Найдем сначала *частное решение* этой системы

$$\mathbf{E} = \mathbf{A}_1 e^{2i(\omega t-\mathbf{kr})}, \quad \mathbf{H} = \mathbf{B}_1 e^{2i(\omega t-\mathbf{kr})},$$

соответствующее *вынужденным колебаниям* с частотой 2ω. Из второго уравнения обычным путем находим, что векторы **E** и **H** *взаимно перпендикулярны*. Аналогично, из последних двух уравнений следует, что $(\mathbf{kA}_1) = (\mathbf{kB}_1) = 0$, т.е. рассматриваемая плоская волна *поперечна как в отношении вектора* **E**, *так и в отношении вектора* **H**. Учтя это, а также соотношение $k^2c^2 = \omega^2\varepsilon(\omega)$, из первых двух уравнений получим

$$\mathbf{A}_1 = \frac{2\pi\alpha_2}{\varepsilon(\omega) - \varepsilon(2\omega)}A\mathbf{A}.$$

Надо еще удовлетворить условию, чтобы на входе в нелинейную среду (где мы поместим начало координат) интенсивность второй гармоники обращалась в нуль. Для этого к частному решению, найденному выше, надо добавить *общее решение* соответствующей однородной системы уравнений и подобрать амплитуду его так, чтобы указанное условие выполнялось. Возвращаясь снова к вещественной форме записи, таким путем получим

$$\mathbf{E}_1 = \frac{2\pi\alpha_2}{\varepsilon(\omega) - \varepsilon(2\omega)}A\mathbf{A}[\cos(\omega_2 t - 2\mathbf{kr}) - \cos(\omega_2 t - \mathbf{k}_2\mathbf{r})], \tag{124.3}$$

где

$$k_2^2 = \omega_2^2\varepsilon(\omega_2)/c^2.$$

3. Таким образом, вторая гармоника представляет собой наложение двух волн одной и той же частоты $\omega_2 = 2\omega$: *вынужденной волны* $\cos(\omega_2 t - 2\mathbf{kr})$ и *свободно распространяющейся волны* $-\cos(\omega_2 t - \mathbf{k}_2\mathbf{r})$. Обе волны распространяются в одном и том же направлении, но *с различными фазовыми скоростями*. Поэтому по мере распространения будет меняться разность фаз между ними и возникнет характерное в таких случаях явление *биений.* Интенсивность I_1 второй гармоники найдется возведением в квадрат и последующим усреднением по времени выражения (124.3). Опуская числовые коэффициенты и обозначая через I интенсивность исходной волны, таким путем найдем

$$I_1 \sim \frac{\alpha_2^2\omega^2x^2I^2}{n^2c^2}\left(\frac{\sin\beta}{\beta}\right)^2, \tag{124.4}$$

где x — расстояние, пройденное волной, и введено обозначение

$$\beta = \frac{(2\mathbf{k}-\mathbf{k}_2)\mathbf{r}}{2} = \frac{(2k-k_2)x}{2}. \tag{124.5}$$

При этом в знаменателе формулы (124.4) мы пренебрегли различием между показателями преломления $n(\omega)$ и $n(2\omega)$.

Когда $\beta = 0,\ \pi,\ 2\pi,\ \ldots$, интенсивность первой гармоники обращается в нуль. Максимумы интенсивности получаются примерно посередине между минимумами. Таким образом, с возрастанием x интенсивность второй гармоники возрастает, когда β лежит приблизительно между нулем и $\pi/2$, между π и $3\pi/2$ и т.д. В этих случаях *энергия переходит от исходной волны ко второй гармонике.* Если же β лежит приблизительно между $\pi/2$ и π, между $3\pi/2$ и 2π и т.д., то с возрастанием x интенсивность второй гармоники убывает. В этих случаях *энергия снова возвращается от второй гармоники к исходной волне.* Такой процесс *перекачки энергии* периодически повторяется по мере распространения исходной волны.

Условие $\beta = \pi/2$ определяет расстояние x, до которого происходит перекачка энергии от исходной волны ко второй гармонике с последующим возвращением ее опять в исходную волну. Это расстояние называется *когерентной длиной.* Для нее из указанного условия нетрудно получить

$$l_{\text{ког}} = \frac{\lambda}{4|n(\omega)-n(2\omega)|}, \tag{124.6}$$

где λ — длина исходной волны, $n(\omega)$ — ее показатель преломления, а $n(2\omega)$ — показатель преломления второй гармоники. *Чем больше когерентная длина, тем интенсивнее происходит перекачка энергии от исходной волны во вторую гармонику.*

Когда

$$n(\omega) = n(2\omega), \quad \text{а потому} \quad 2\mathbf{k} = \mathbf{k}_2, \tag{124.7}$$

то длина когерентности $l_{\text{ког}}$ обращается в бесконечность. В этом случае переход энергии особенно интенсивен и должен происходить от исходной волны к ее второй гармонике. Обе волны $\cos(\omega_2 t - 2\mathbf{k}\mathbf{r})$ и $-\cos(\omega_2 t - \mathbf{k}_2\mathbf{r})$ распространяются с одинаковыми фазовыми скоростями, а потому фазовое соотношение между ними сохраняется все время при их распространении. С этим, как и при всяком резонансе, и связана эффективность обмена энергией между взаимодействующими волнами. Поэтому условие (124.7) называется *фазовым синхронизмом* между рассматриваемыми волнами. В этом случае $\beta = 0$ при любом x, и из (124.4) предельным переходом $\beta \to 0$ получаем

$$I_1 = \frac{\alpha_2^2 \omega^2 x^2 I^2}{n^2 c^2}. \tag{124.8}$$

Конечно, при достаточно больших x формула (124.8) перестает быть верной, так как при $x \to \infty$ она дает $I_1 \to \infty$, а интенсивность второй гармоники не может превосходить интенсивности I исходной волны. В этом случае метод последовательных приближений, с помощью которого была получена формула (124.8), неприменим.

Когерентную длину $l_{\text{ког}}$ можно определить как максимальное расстояние, на котором приближенно сохраняется фазовый синхронизм между падающей и переизлученной волнами, т.е. расстояние, на котором набег фазы одной из этих волн относительно другой не превышает π.

4. Генерация второй гармоники впервые была обнаружена Франкеном в 1961 г. Излучение рубинового лазера фокусировалось на кристалле кварца, расположенном между двумя фильтрами, прозрачными только в узких окрестностях длин волн $\lambda_1 = 694{,}3$ нм и $\lambda_2 = \lambda_1/2 = 347{,}0$ нм. Из кристалла, помимо исходного красного излучения лазера ($\lambda_1 = 694{,}3$ нм), выходило также ультрафиолетовое излучение с половинной длиной волны 347,0 нм. В первых опытах Франкена в энергию второй гармоники превращалось $\sim 10^{-8}$ энергии первичного излучения.

Почему лишь столь ничтожная доля энергии переходила ко второй гармонике? Это объясняется *малостью когерентной длины* $l_{\text{ког}}$ в кварце. Для интенсивного обмена энергией надо удовлетворить *условию фазового синхронизма* $n(\omega) = n(2\omega)$. Но это невозможно сделать для *изотропных сред в прозрачной области* спектра, так как в этой области показатель преломления $n(\omega)$ *монотонно возрастает с частотой.* Равенство $n(\omega) = n(2\omega)$ может удовлетворяться только тогда, когда частота ω взята в прозрачной области, а 2ω — в области сильного поглощения или наоборот.

Однако, как указали в 1962 г. Джордмейн и Терхьюн и подтвердили это указание на опыте, *фазовый синхронизм на частотах ω и 2ω можно осуществить между обыкновенной и необыкновенной волнами в некоторых кристаллах.* На рис. 353 представлены обыкновенный n^o и необыкновенный n^e показатели преломления одноосного кристалла в зависимости от направления волновой нормали. Сплошные кривые относятся к частоте ω, штриховые — к удвоенной частоте 2ω. На первом из этих двух рисунков кривые $n^o(\omega)$ и $n^e(2\omega)$ пересекаются между собой. Точкам их пересечения соответствуют направления, для которых между обыкновенной волной с частотой ω и ее гармоникой с частотой 2ω выполняется фазовый синхронизм. Эти направления называются *направлениями синхронизма*, а угол ϑ между ними и оптической осью кристалла — *углом синхронизма*.

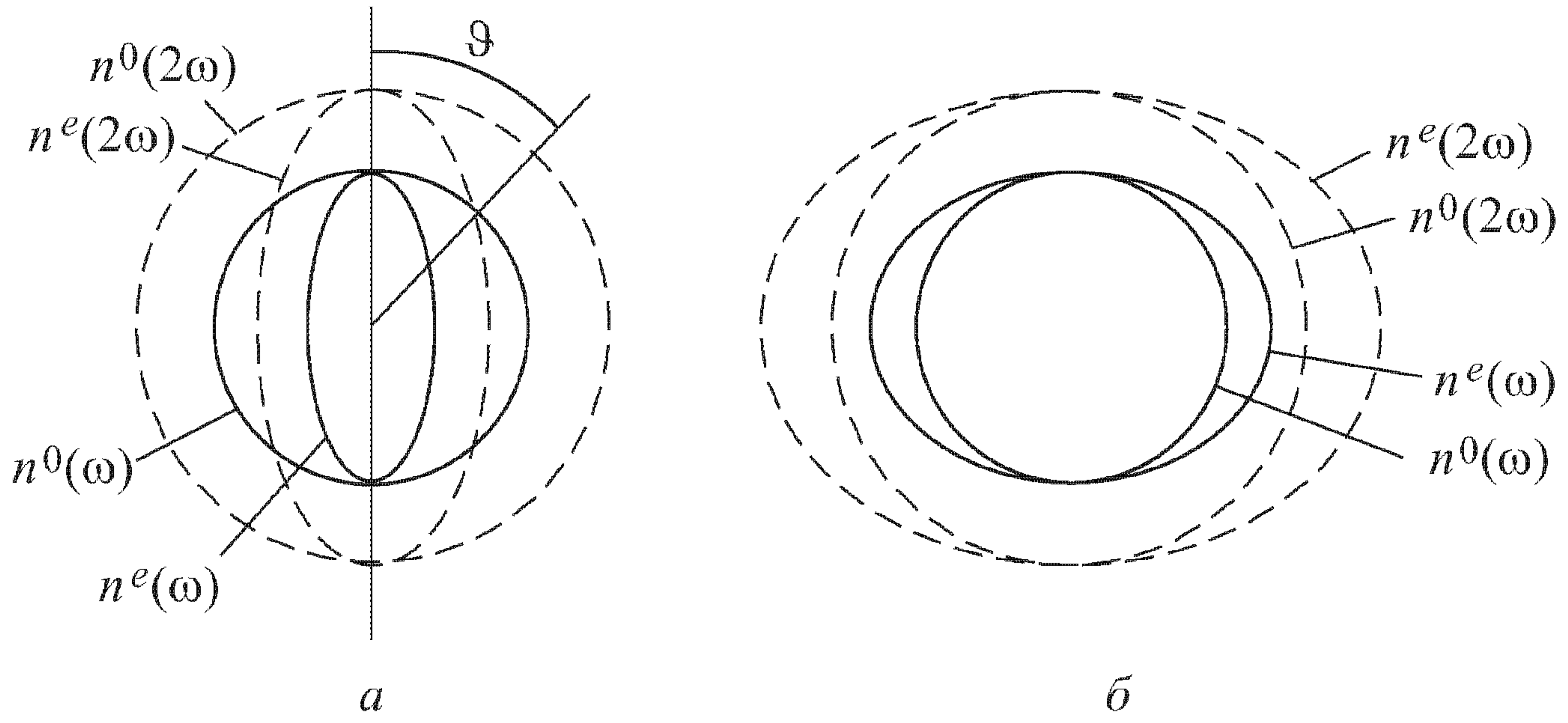


Рис. 353

Хотя обе волны и поляризованы в различных плоскостях, но они могут нелинейно взаимодействовать между собой, поскольку в кристаллах квадратичная поляризуемость есть не скаляр α_2, а *тензор* $(\alpha_2)_{jkl}$. Поэтому при наличии фазового синхронизма должна происходить перекачка энергии от исходной волны к ее второй гармонике, что и наблюдается на самом деле. Таким путем удается более половины падающего света превратить во вторую гармонику. Понятно, что при таких больших интенсивностях второй гармоники метод последовательных приближений может оказаться неточным и даже неприменимым. Однако качественное заключение о влиянии фазового синхронизма *остается в силе*.

Подходящим кристаллом может быть одноосный кристалл *дигидрофосфата калия* KH_2PO_4 (сокращенно KDP). Для этого кристалла при $\lambda = 1{,}15$ мкм, как показывает расчет, подтверждаемый наблюдениями, угол синхронизма равен $41^\circ 35'$. Существует красивый демонстрационный опыт. Кристалл KDP, вырезанный параллельно оптической оси, кладется на столик, который может вращаться вокруг вертикальной оси. Оптическая ось кристалла должна быть горизонтальна. На кристалл направляется мощный инфракрасный луч от лазера на неодимовом стекле ($\lambda = 1060$ нм). Луч лазера невидим, но его можно обнаружить с помощью листа черной бумаги. Бумага загорается, если ее поместить на пути луча. При произвольной ориентации кристалла никакого видимого света не возникает. Но если кристалл медленно поворачивать, то из него выходит ослепительно яркий зеленый луч ($\lambda = 530$ нм), когда станет выполняться условие синхронизма.

Не всякий одноосный кристалл годится для опыта такого типа. Не годится, например, кристалл кварца, в котором, как видно из рис. 353 *б*, условие фазового синхронизма не выполняется ни для какого направления.

5. Остановимся теперь на *генерации волн с суммарной и разностной частотами.* Природа этого явления в точности такая же, что и генерация второй гармоники. Поэтому достаточно только указать, в чем состоит явление. Если на нелинейную среду направить два мощных пучка света с различными частотами ω_1 и ω_2, то из нее будет выходить свет *не только с первоначальными частотами* ω_1 и ω_2 *и их гармониками* $2\omega_1$ *и* $2\omega_2$, *но и свет с суммарной* $\omega_1 + \omega_2$ *и разностной* $\omega_1 - \omega_2$ *частотами.* Подобными методами генерации волн разных частот удается далеко проникнуть в инфракрасную и ультрафиолетовую области спектра. Например, удалось получить ультрафиолетовое излучение с длиной волны ≈ 80 нм.

§ 125. Второе приближение. Самофокусировка

1. Для нахождения *второго приближения* надо использовать вектор $\mathbf{P}_{\text{нл}}$ в первом приближении, т.е.

$$\mathbf{P}_{\text{нл}} = \alpha_2(E_0 + E_1)(\mathbf{E}_0 + \mathbf{E}_1) + \alpha_3 E_0^2 \mathbf{E}_0.$$

Однако мы ограничимся *только изотропными средами или кристаллами, обладающими центром симметрии.* Для них, как было показано в § 123, $\alpha_2 = 0$, и следовательно, в нужном приближении $\mathbf{P}_{\text{нл}} = \alpha_3 E_0^2 \mathbf{E}_0$. Подставив сюда выражение (123.5), получим

$$\mathbf{P}_{\text{нл}} = \frac{3\alpha_3 A^2}{4}\mathbf{A}\cos(\omega t - \mathbf{kr}) + \frac{\alpha_3 A^2}{4}\mathbf{A}\cos[3(\omega t - \mathbf{kr})]. \tag{125.1}$$

Слагаемое с тройной частотой $\omega_3 = 3\omega$, очевидно, приводит к *генерации третьей гармоники.* Разность показателей преломления $n(3\omega) - n(\omega)$ здесь еще больше, чем в случае второй гармоники. Это ограничивает выбор кристаллов, для которых можно удовлетворить *условию фазового синхронизма.* Основная трудность опыта связана с малыми значениями кубичной поляризуемости α_3, что вынуждает применять большие кристаллы и большие освещенности, часто приводящие к разрушению кристаллов. Несмотря на это, генерация третьей гармоники наблюдалась еще в 1962 г. группой американских физиков на кристалле исландского шпата при освещении его светом рубинового лазера. На выходе кристалла удалось зарегистрировать излучение с длиной волны 231,3 нм. Позднее генерация наблюдалась в некоторых оптически изотропных кристаллах (например, LiF, NaCl), жидкостях и газах.

В третьем приближении, помимо гармоник более низкого порядка, очевидно, должна появиться *четвертая*, а в следующем приближении — *пятая* гармоники и т.д. Экспериментально четвертая гармоника наблюдалась С.А. Ахмановым в 1974 г., а пятая — Харрисом в 1973 г.

2. Посмотрим теперь, какие явления связаны с первым слагаемым в выражении (125.1). Множитель $\mathbf{A}\cos(\omega t-\mathbf{kr})$ есть исходная падающая волна $\mathbf{E}_0$. Ясно, что в рассматриваемом приближении его можно заменить на $\mathbf{E}$, так как такая замена сказывается лишь на членах *более высокого порядка малости*, не учитываемых в рассматриваемом приближении. После этого (123.4) запишется в виде

$$\operatorname{rot}\mathbf{H} - \frac{1}{c}[\varepsilon(\omega) + 3\pi\alpha_3(\omega)A^2]\frac{\partial\mathbf{E}}{\partial t} = 0.$$

Отсюда видно, что влияние рассматриваемого слагаемого эквивалентно *изменению диэлектрической проницаемости или показателя преломления среды.* Учитывая малость поправки к $\varepsilon(\omega)$, для показателя преломления n в поле интенсивной световой волны можно написать

$$n = n_0 + n_2A^2, \tag{125.2}$$

где n_0 — значение показателя преломления среды в линейной оптике, а $n_2(\omega)$ — некоторый коэффициент, зависящий от свойств среды.

Помимо рассмотренной, есть и другие причины изменения показателя преломления в электрическом поле. В нелинейной среде из-за *электрострикции* световая волна вызывает появление *постоянного давления*, аналогично тому как появляется постоянное слагаемое в формуле (124.1). Это приводит *к изменению плотности и показателя преломления среды.* В жидкостях с анизотропными молекулами показатель преломления изменяется из-за *высокочастотного эффекта Керра* (см. § 90). Показатель преломления всегда изменяется из-за *нагревания среды световой волной.* Во всех этих случаях изменение показателя преломления пропорционально *квадрату амплитуды*, а потому может быть также представлено формулой (125.2).

Постоянная n_2 может быть и отрицательной, и положительной. Она особенно велика у нитробензола и имеет для него положительный знак.

3. Из изложенного следует, что если через однородную среду проходит интенсивный пучок света, то среда становится *оптиче-*

ски неоднородной. Луч света в такой среде загибается в сторону большего показателя преломления. С этим связано явление *самофокусировки* (когда $n_2 > 0$) и *дефокусировки* (когда $n_2 < 0$) света, предсказанное теоретически Г.А. Аскарьяном в 1962 г. и впервые наблюдавшееся Н.Ф. Пилипецким и А.Р. Рустамовым в 1965 г. Затем самофокусировка наблюдалась для многих газов, жидкостей и твердых тел.

Чтобы простейшим путем понять сущность явления самофокусировки, предположим, что в однородную среду с показателем преломления n_0 вступает плоскопараллельный пучок лучей кругового поперечного сечения с диаметром D (рис. 354). Допустим сначала, что амплитуда пучка постоянна по всему сечению. Показатель преломления в пространстве, занятом пучком, сделается равным $n = n_0 + n_2A^2$, причем мы предположим, что $n_2 > 0$. Из-за дифракции пучок расширяется. Практически все направления лучей сосредоточатся в пределах конуса с углом при вершине $2\vartheta_{\text{диф}}$, где $\vartheta_{\text{диф}} = 1{,}22\lambda/(Dn_0)$, а λ — длина волны в вакууме. (Направления лучей относятся к пространству внутри цилиндра.) Предельный угол скольжения ϑ_0 для полного отражения от боковой стенки цилиндра определяется соотношением

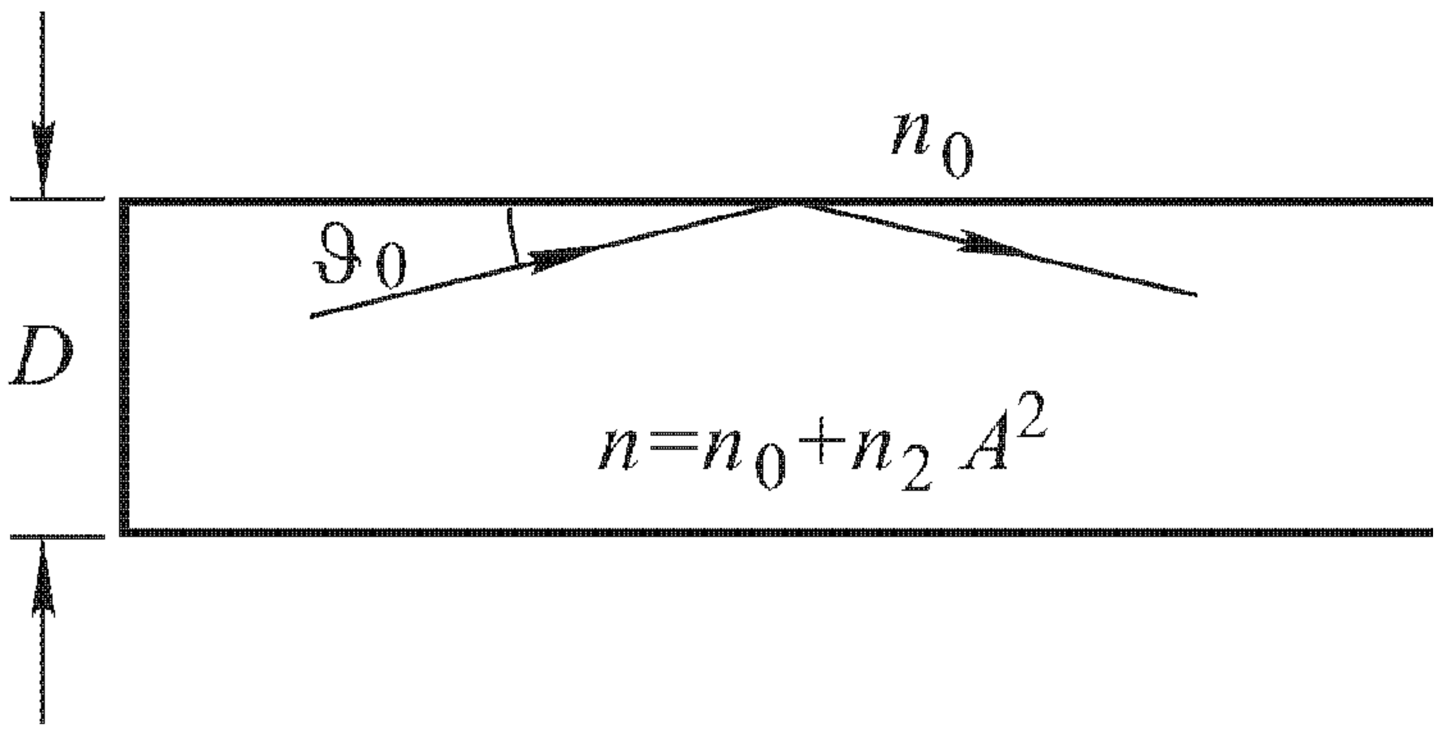


Рис. 354

$$\cos\vartheta_0 = n_0/(n_0 + n_2A^2).$$

Ввиду малости этого угла отсюда находим: $1 - \cos\vartheta_0 \approx A^2n_2/n_0$ и, следовательно,

$$\vartheta_0^2 \approx 2A^2n_2/n_0.$$

Если $\vartheta_{\text{диф}} > \vartheta_0$, то часть дифрагированных лучей будет выходить из цилиндра — пучок будет *расширяться.* При обратном соотношении $\vartheta_{\text{диф}} < \vartheta_0$ все дифрагированные лучи будут испытывать полное отражение от боковой поверхности цилиндра. А так как в реальных условиях интенсивность света и показатель преломления возрастают к оси пучка, то из-за искривления лучей пучок начнет *сжиматься и может стянуться в тонкий шнур.* Это и есть *самофокусировка.*

В промежуточном случае, когда $\vartheta_{\text{диф}} \approx \vartheta_0$, пучок будет проходить через нелинейную среду практически без изменения поперечных размеров. Он создает для себя как бы волновод, в котором и распространяется без рассеяния в стороны. Такой режим распространения называется *самоканализацией светового пучка.* Таким образом, самоканализация имеет место при условии $\vartheta_0 \approx \vartheta_{\text{диф}}$. Подставив сюда значения углов ϑ_0 и $\vartheta_{\text{диф}}$, а также выражение амплитуды A через мощность пучка

$$P = \frac{cn_0 A^2}{8\pi}\frac{\pi D^2}{4} = \frac{cn_0 D^2}{32}A^2,$$

получим так называемую *пороговую мощность*, выше которой начинается сжатие пучка. Она определяется соотношением

$$P_{\text{порог}} \approx c\frac{(0{,}61\lambda)^2}{16 n_2}. \tag{125.3}$$

Расстояние от края среды, на котором фокусируются крайние лучи пучка, легко оценить из следующих соображений. В пучке угловое расхождение лучей из-за дифракции равно $2\vartheta_{\text{диф}}$. При критической мощности в результате отражения от боковой поверхности пучка крайние лучи делаются параллельными. Это произойдет на расстоянии

$$f_{\text{эф}} = \frac{D}{2\vartheta_{\text{диф}}} \approx \frac{n_0 D^2}{2{,}44\lambda}. \tag{125.4}$$

Оно играет при самофокусировке роль *эффективного фокусного расстояния для крайних лучей пучка.* Если вместо расходящихся лучей взять лучи, параллельные оси пучка, то они сфокусируются на том же расстоянии $f_{\text{эф}}$.

Для сероуглерода CS_2, обладающего сравнительно большим значением $n_2 = 2 \cdot 10^{-11}$ СГСЭ, при освещении рубиновым лазером ($\lambda = 694{,}3$ нм) пороговая мощность, вычисленная по формуле (125.3), равна $P_{\text{порог}} \approx 17$ кВт. Если диаметр пучка $D = 1$ мм, то формула (125.4) в этом случае дает $f_{\text{эф}} \approx 96$ см ($n_0 = 1{,}62$). В некоторых сортах оптического стекла $P_{\text{порог}} \sim 1$ Вт. В этих случаях явление самофокусировки можно наблюдать не только в мощных пучках импульсных лазеров, но и в малоинтенсивных пучках лазеров непрерывного действия.

§ 126. Параметрическая генерация света

1. Нелинейные оптические явления в кристаллах позволяют преобразовывать излучение лазера не только в излучения гармоник, суммарных и разностных частот, но и в *излучения с плавно перестраиваемой частотой.* Принцип такого преобразования был указан в 1962 г. С.А. Ахмановым и Р.В. Хохловым (1926–1977). Он заключается в следующем. Пусть на среду, нелинейная поляризация которой с точностью до квадратичных членов определяется выражением $\mathbf{P}_{\text{нл}} = \alpha_2 E\mathbf{E}$, падает мощная «волна накачки» $\mathbf{E}_{\text{н}} = \mathbf{A}_{\text{н}}\cos(\omega_{\text{н}}t - \mathbf{k}_{\text{н}}\mathbf{r})$ и две слабые волны $\mathbf{E}_1 = \mathbf{A}_1\cos(\omega_1 t - \mathbf{k}_1\mathbf{r})$ и $\mathbf{E}_2 = \mathbf{A}_2\cos(\omega_2 t - \mathbf{k}_2\mathbf{r})$, частоты которых связаны соотношением

$$\omega_{\text{н}} = \omega_1 + \omega_2. \tag{126.1}$$

Считая для простоты, что направления амплитуд всех волн совпадают, перейдем к скалярной форме записи. В первом приближении нелинейная поляризация среды будет равна $\alpha_2(E_{\text{н}} + E_1 + E_2)^2$. Возведя в квадрат, рассмотрим член $2\alpha_2 E_1 E_{\text{н}}$, представляющий собой произведение двух косинусов. Преобразуем его в сумму двух косинусов и возьмем слагаемое с разностной частотой $(\omega_{\text{н}} - \omega_1)$, которая, ввиду (126.1), равна ω_2. Так же поступим с произведением $2\alpha_2 E_2 E_{\text{н}}$. В результате из нелинейной поляризации $P_{\text{нл}}$ выделятся два члена с частотами ω_1 и ω_2:

$$\begin{aligned} P_{\text{нл}}(\omega_1) &= \alpha_2 A_2 A_{\text{н}} \cos\left[\omega_1 t - (\mathbf{k}_{\text{н}} - \mathbf{k}_2)\mathbf{r}\right], \\ P_{\text{нл}}(\omega_2) &= \alpha_2 A_1 A_{\text{н}} \cos\left[\omega_2 t - (\mathbf{k}_{\text{н}} - \mathbf{k}_1)\mathbf{r}\right]. \end{aligned} \tag{126.2}$$

Следовательно, возникнет переизлучение волн с теми же частотами ω_1 и ω_2. Это может привести к *усилению волн таких частот* за счет энергии волны накачки. Такое явление называется *параметрическим усилением света*, так как его можно рассматривать как результат *модуляции параметров среды* (показателя преломления) при ее взаимодействии с волной накачки. Оно было открыто в 1965 г. Ахмановым и Хохловым с сотрудниками в СССР, а также Джордмейном и Миллером в США и использовано ими для создания *когерентных генераторов света с плавно перестраиваемой частотой.* Взаимодействие с волной накачки будет особенно сильным, когда фазы волн (126.2) длительно совпадают с фазами обеих волн E_1 и E_2, т.е. когда соблюдается условие

$$\mathbf{k}_1(\omega_1) + \mathbf{k}_2(\omega_2) = \mathbf{k}_{\text{н}}(\omega_{\text{н}}). \tag{126.3}$$

Это условие называется условием *фазового синхронизма* между волной накачки и обеими волнами с частотами ω_1 и ω_2. Полученное ранее условие (124.7) является частным случаем условия (126.3). Чтобы в этом убедиться, достаточно записать (124.7) в виде

$$\mathbf{k}(\omega) + \mathbf{k}(\omega) = \mathbf{k}_2(2\omega)$$

и применить его к процессу образования волны с частотой ω из ее второй гармоники.

Если условие синхронизма выполнено, то энергия от волны накачки будет в нелинейной среде передаваться волнам с частотами ω_1 и ω_2. Для эффективного усиления этих волн надо волну накачки заставить многократно проходить через нелинейную среду (кристалл). Для этого последнюю, как в лазерах, помещают в *оптический резонатор между двумя зеркалами.* Оба зеркала должны иметь достаточно высокие коэффициенты отражения для волн обеих частот ω_1 и ω_2 и в то же время одно из них, через которое входит волна накачки, должно быть в достаточной степени прозрачным для этой волны. При достаточно высоких коэффициентах отражения зеркал и большой мощности волны накачки возникает генерация на частотах ω_1 и ω_2, удовлетворяющих условиям (126.1) и (126.3).

Нет необходимости специально посылать в резонатор волны с частотами ω_1 и ω_2. Они сами возникают либо из-за всегда имеющихся *шумов*, либо из-за *тепловых флуктуаций*. Происходит *самовозбуждение* генератора с последующим усилением генерируемых волн при нелинейном взаимодействии их с волной накачки. В качестве волны накачки обычно используется вторая (или третья) гармоника рубинового или неодимового лазера.

2. *В изотропных средах в области нормальной дисперсии нельзя удовлетворить одновременно обоим условиям* (126.1) *и* (126.3). Действительно, допустим сначала, что все три волны с частотами $\omega_\text{н}$, ω_1, ω_2 распространяются в одном направлении. В этом случае условие (126.3) можно записать в виде $\omega_1 n_1 + \omega_2 n_2 = \omega_\text{н} n_\text{н}$, где n_1, n_2, $n_\text{н}$ — показатели преломления для соответствующих частот. С учетом условия (126.1) отсюда получаем $(n_\text{н} - n_1)\omega_1 + (n_\text{н} - n_2)\omega_2 = 0$, а это невозможно, так как $(n_\text{н} - n_1) > 0$ и $(n_\text{н} - n_2) > 0$. Из приведенного рассуждения следует, что волновое число $k_\text{н}$ всегда больше суммы волновых чисел k_1 и k_2, *независимо от направления волн*. Поэтому условию (126.3) нельзя удовлетворить и при различных направлениях векторов $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_\text{н}$. Иначе получился бы векторный треугольник, одна сторона которого длиннее суммы длин двух других сторон.

Однако *синхронизм можно получить в некоторых кристаллах между обыкновенной и необыкновенной волнами*. Только теперь, когда частоты ω_1 и ω_2 могут не совпадать, для осуществления синхронизма имеется больше возможностей, чем в аналогичном случае при генерации второй или третьей гармоник (см. § 124). В принципе синхронизм мог бы осуществляться в четырех случаях:

$$\begin{aligned} &1)\ \mathbf{k}_\text{н}^e = \mathbf{k}_1^o + \mathbf{k}_2^e, \quad 2)\ \mathbf{k}_\text{н}^e = \mathbf{k}_1^o + \mathbf{k}_2^o, \\ &3)\ \mathbf{k}_\text{н}^o = \mathbf{k}_1^o + \mathbf{k}_2^e, \quad 4)\ \mathbf{k}_\text{н}^o = \mathbf{k}_1^e + \mathbf{k}_2^e, \end{aligned} \qquad (126.4)$$

где индексы o и e относятся к обыкновенной и необыкновенной волнам. Разумеется, не всем этим условиям, и даже хотя бы одному из них, можно удовлетворить в реальных кристаллах. Так, в случае одноосного кристалла дифосфата калия KH_2PO_4 можно удовлетворить первым двум условиям. В первом случае в «направлении синхронизма» необыкновенная волна накачки будет генерировать в кристалле обыкновенную и необыкновенную волны с частотами ω_1 и ω_2, а во втором случае обе генерируемые волны будут обыкновенными. Для одноосного кристалла ниобата лития $LiNbO_3$, обладающего очень большой нелинейностью, можно удовлетворить только второму из условий (126.4).

Поворачивая кристалл (или изменяя его температуру, а также накладывая постоянное электрическое поле), можно изменять частоты ω_1 и ω_2, для которых направление, перпендикулярное к зеркалам, является направлением синхронизма. Именно так действуют *параметрические генераторы когерентного света, позволяющие плавно перестраивать частоту.* КПД таких генераторов, определяемый как отношение мощностей параметрически генерируемых волн к мощности волны накачки, достигает нескольких процентов при выходной мощности в несколько десятков и сотен киловатт. Ясно, что генерируемые частоты ω_1 и ω_2 *всегда меньше* частоты $\omega_{\text{н}}$ волны накачки. Плавно перестраивая параметрический генератор света, можно пройти весь диапазон видимого света от красного до фиолетового, а также далеко проникнуть и в инфракрасную область спектра.

3. Если (126.1) умножить на постоянную Планка $\hbar$, то получится

$$\hbar\omega_{\text{н}} = \hbar\omega_1 + \hbar\omega_2. \tag{126.5}$$

Это соотношение в квантовой физике интерпретируется как процесс *распада фотона $\hbar\omega_{\text{н}}$ на два фотона $\hbar\omega_1$* и $\hbar\omega_2$, причем уравнение (126.5) выражает *закон сохранения энергии* для этого элементарного акта. Аналогично, генерация второй гармоники с квантовой точки зрения есть процесс *взаимодействия двух фотонов* с энергией $\hbar\omega$ каждый, в результате которого *рождается фотон $\hbar\omega_2$ с удвоенной частотой $\omega_2 = 2\omega$*. Точно так же можно интерпретировать генерацию третьей и высших гармоник, а также генерацию волн с суммарной и разностной частотами. Однако сейчас мы не будем входить в обсуждение всех этих вопросов, так как квантовые явления предполагается рассмотреть в пятом томе нашего курса.

ИМЕННОЙ УКАЗАТЕЛЬ

Абе 115, 118, 122–124, 126, 127, 129, 166, 175–177, 386, 388, 393–395
Абрагам 590
Авогадро 553, 742
Айвс 268, 691, 692
Амичи 126, 127
Ангерер 289
Андерсон 334
Араго 28, 203, 288, 289, 508, 513, 547, 605
Аренс 493, 494
Аркадьев 304
Аскарьян 775
Ахманов 774, 776, 777

Бабине 296, 297, 327, 371, 502–504
Бальмер 603
Бартолинус 487
Басов 751
Беккерель 615
Беннет 763
Бессель 95, 314
Бийе 213, 215, 220, 221
Био 28, 605, 607
Бозе 742, 743
Больцман 150, 563, 589, 723–726, 728, 730, 731, 738, 740–742, 744, 745
Бонч-Бруевич 666
Бор 32, 743, 745
Борн 555, 590
Брадлей 692, 694
Браун 407
Бреге 29
Бриллюэн 66, 643, 645, 646, 648–650, 766
Бродхун 161
Бройль де 32, 45, 387
Брэгг 413–416, 418, 644, 645
Брюстер 28, 432, 433, 435, 446–448, 455, 456, 459, 460, 463–465, 476, 518, 759
Бугер 582, 583

Вавилов 151, 271–276, 511, 582, 583
Ван-Циттер 240, 242
Вебер 30, 40
Ведам 625
Вейерштрасс 124, 126
Верде 612, 615
Вильямс 337
Вин 578–580, 616, 617, 726, 727, 729, 730, 732, 735, 737, 738, 740, 742, 745, 747
Винер 267–269
Владимирский 648
Волластон 495, 497
Вольфке 363
Вуд 287, 334, 357, 476, 616
Вульф 413–416, 418, 644, 645

Габор 363–365
Гаген 477
Гайдингер 611
Гайзенберг 229
Галилей 184, 656, 657, 659, 676, 703
Гамильтон 129, 541, 543, 736
Гартнак 493, 494
Гаусс 78
Гельмгольц 77, 83, 85, 124, 132, 147, 388, 446
Герке 265, 266, 337–339, 342, 343
Герц 30, 314, 450, 455, 530, 633, 643, 768
Гершель 131, 132, 186–187
Гинзбург 273, 276
Глазебрук 493, 494, 497
Глан 493, 494
Горелик 209
Грегори 187, 190
Гримальди 20, 25, 211, 222
Грин 305
Гросс 527, 625, 646
Гук 20, 25
Гульстранд 143, 144
Гюйгенс 20, 21, 23–26, 47–49, 178-181, 183, 238, 242, 277–279, 281, 282, 291, 292, 294, 304, 305, 307, 309, 325, 335, 354, 374, 380, 393, 404, 487, 537

Дебай 417, 555, 646, 732
Декарт 14, 15, 20, 25
Делиль 288
Денисюк 370, 373, 374
Джаван 663, 763
Джасея 663
Джермер 32
Джинс 731, 735–740, 745, 747, 762

Джордмейн 771, 777
Доплер 426, 453, 582, 688–692, 695, 696, 698, 725, 726
Друде 268, 461, 463, 464, 473, 475, 477, 607
Дэвиссон 32

Жамен 249, 251, 252, 254, 256, 565, 584

Зееман 31, 210, 596, 597, 601, 602, 604, 613, 616, 706
Зельмейер 551, 569, 575, 614
Зоммерфельд 66, 314

Ингл 353

Капица 599, 601
Каплянский 527, 625
Каратеодори 134, 136
Карпо 725
Кассегрен 187, 190
Кельнер 180
Кеннеди 664
Кеплер 183, 184, 559, 666
Керр 31, 200, 583–586, 588–594, 596, 759, 774
Кеттелер 471, 474
Кирхгоф 277, 279, 281, 292, 304, 307, 717, 719 723, 735
Китинг 686
Клаузиус 46, 554, 721, 722, 735
Книппинг 407
Кольрауш 31, 40
Комптон 32
Коперник 692
Корбино 615
Корню 298–303
Коттон 31, 593, 594, 611
Кришнан 650
Кун 562
Кундт 564, 565, 612, 613
Курльбаум 737

Лагранж 77, 83, 85, 124, 132
Ладенбург 563
Ламберт 159, 160, 162, 163, 165, 398, 720, 721
Ламе 530
Ландау 646, 647
Ландсберг 642, 646, 649, 651
Ланжевен 586, 589, 593
Лаплас 44
Лармор 655, 675
Лауэ 407, 410 412, 414, 416
Лебедев 559, 724
Левин 573, 575
Лейт 363
Лемуан 590
Леонтович 646
Леру 564
Липпман 269–271, 373
Ллойд 213, 220–222, 234, 243, 541, 543
Ломоносов 28, 169, 186
Лорентц 31, 527, 548, 554, 555, 580, 596, 598, 602, 613, 655, 659, 675–678, 680, 681, 689, 703, 705–708, 710
Лоренц 554, 555
Лошмидт 639
Луммер 161, 265, 266, 337–339, 342, 343, 737
Лэнглей 737

Майкельсон 30, 223, 240, 256–258, 333, 334, 337, 341, 362, 404–407, 659–664, 669, 681, 705
Макалюзо 615
Максвелл 29, 30, 38, 39–41, 44, 45, 137, 139, 292, 314, 424, 425, 453, 461, 466, 478, 481, 485, 519, 548, 551, 581, 609, 622, 632, 633, 637, 659, 665, 707, 724, 767
Максутов 188–190
Малюс 28, 56, 57, 422, 433
Мандельштам 395, 579, 584, 639, 643, 645, 646, 648–651, 766
Маральди 288
Марци 20
Мах 256, 275
Мейман 754, 764
Меслин 215, 220
Миллер 777
Минковский 677–680, 707, 708
Михельсон 731
Молчанов 666
Морзе 315
Морли 257, 662
Мосотти 554
Муррей 663
Мутон 31, 593, 594

Нейхаусс 271
Нернст 268
Николь 491–495, 497
Ньютон 20, 21, 25, 27–29, 33, 84, 172, 186, 190, 247, 248, 437, 548, 549, 564, 656, 707

Оршштейн 641

Пастернак 625
Паули 665
Пашен 353, 737

Перо 263–265, 267, 337–339, 350, 751–753
Пилипецкий 775
Пифагор 20
Планк 10, 31, 387, 577, 715, 736–740, 742, 743, 745, 747, 779
Плачек 646, 647
Поггендорф 541
Пойнтинг 434, 436, 482, 501–503, 530, 532, 576, 634
Поккельс 594, 596, 759
Поль 214, 215, 220, 221, 289, 290, 398
Празмовский 493, 494
Прево 718, 719
Принсхейм 737
Прохоров 751
Пуанкаре 655, 675
Пуассон 288, 289, 725
Пуркинье 148
Пуччианти 565

Раман 650
Рамсден 95, 178, 180, 181, 183
Резерфорд 333
Рейхе 562
Рёмер 20, 694
Риттенгауз 333
Риттер 493, 494
Ритц 665
Рождественский 256, 388, 395, 565, 566, 569
Роулэнд 334, 350, 352, 353
Рошон 495
Рубенс 476, 477, 737
Рубинштейн 315
Рустамов 775
Рытов 325
Рэлей 60, 211, 278, 287, 314, 332, 336, 338, 341, 354, 357–359, 363, 366, 379–381, 384, 386, 390, 395, 451, 463, 528, 551, 576, 617, 632, 636, 639, 641, 731, 735–740, 745, 747, 762

Сенармон 495
Сивухин 45
Ситтер де 666
Смолуховский 637, 639, 641
Снеллиус 14, 15, 22, 33, 50, 53, 275, 486, 496, 544
Солейль 503, 504
Стефан 150, 723–726, 728, 730, 740–742
Стилуэлл 691, 692
Столетов 433, 448, 449

Тамм 273, 274
Таунс 663, 751
Твисс 407
Терхьюн 771
Тиндаль 632, 636
Томас 562
Томашек 666
Томпсон 493, 494
Торндайк 664

Уайтфорд 303
Угаров 680
Ульянин 721
Умов 532
Упатинекс 363

Фабелинский 325
Фабри 263–265, 267, 337–339, 350, 751–753
Фабрикант 750
Фарадей 29, 31, 553, 583, 600, 605, 611, 612, 614, 742
Ферма 49–54, 56–57, 87, 88, 340
Физо 29, 32, 249, 403, 404, 705
Фицжеральд 681
Форрестор 210
Фохт 541, 542, 603, 615
Франк И.М. 273, 274, 276
Франк (поляризационная призма) 493, 494
Франкен 771
Франклин 28
Фраунгофер 291, 293–296, 309, 315, 317, 320, 333, 378
Френель 28, 29, 203, 211–214, 216, 220, 221, 225, 269, 277–289, 291–294, 296, 299, 304, 307–311, 314, 315, 325, 354, 374, 393, 430–432, 434–436, 439, 440, 443, 446, 447, 449, 451, 454–460, 462, 463, 472, 473, 478, 508, 523, 525, 526, 533, 571, 608–611, 613, 705
Френкель 27
Фридрих 407
Фуко 29, 32, 492–494, 697, 699
Фурье 41, 58, 225–227, 230, 234, 239, 344–346, 355, 357–362, 370, 371, 390, 391, 435, 578, 644, 645, 732

Харрис 774
Хафель 686
Хелл 417
Хохлов 776, 777
Христиансен 564
Хьюиш 572

Цейсс 123, 175, 438
Цендер 256

Цернике 240, 242, 400, 641
Циолковский 711

Черенков 271–276

Шерер 417
Шмидт 188–190
Шредингер 560
Штарк 581, 602–604
Шустер 298, 299, 302

Эйлер 28, 29
Эйнштейн 30, 31, 637, 641, 655, 656, 658, 660, 665, 669, 670, 672, 676, 682, 699, 702, 709, 742–744, 749
Эйри 378–383, 386, 392, 518, 695
Эренфест 60, 61, 736
Эрриот 763

Юнг 28, 29, 147, 210, 211, 222, 248, 403

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Аббе способ испытания объективов 127
— число 115
Аберрации оптических систем 101, 106, 113
Аберрационная кривая 107
Аберрация света 688, 692
Абсолютно матовое тело 721
— черное тело 719
Абсолютный оптический инструмент 136
Адаптация 151
Адиабатические инварианты 725, 728, 729
Аккомодация 141
—, область 141
Активная (инверсно заселенная) среда 750
Анаберрационная поверхность 55
Анализатор 421
Апертура 96
— полной поляризации 492
— числовая 166
Апертурный угол 97
Апохромат 118, 175
Астигматизм 101, 110
Астигматическая разность 102
Астрономическая рефракция 34
Атомные плоскости 410
— прямые 410
Ахромат 118, 176
Ахроматизм положения 176
— увеличения 177

Бабине компенсатор 502
— теорема 297
Бабине Солейля компенсатор 503
Баллистическая гипотеза 665
Беккереля формула 615
Бийе билинза 213
Бинокль 184
Бора правило частот 745
Брэгга Вульфа условие 413
Брюстера закон 432, 456, 459, 465
— угол 432
Бугера закон 582

Вектор импульса–энергии 710
Венцы 319
Верде постоянная 612
Видимая величина площадки 153
Видности кривая 148
Вина закон смещения 729, 730
— парадокс 616
— постоянная 742
— теорема 726
Виньетирование 98
Водянистая влага 139
Волна де Бройля 387
— дипольная 455
— монохроматическая 201
— необыкновенная 485
— неоднородная 43
— нормальная 610
— обыкновенная 484
— однородная 43
— поверхностная 43
— поляризации 455
— стоячая световая 267
Волновое сопротивление 274
— уравнение 45
— число спектроскопическое 229
Время восстановления 59, 63
— жизни 578
— затягивания 346
— установления 200
Вторичные волны Гюйгенса 24
Вторичный спектр 113
Входной зрачок 96
— люк 98
Выходной зрачок 97
— люк 98

Галилея преобразование 656
— принцип относительности 657
Гало 319
Гамма-излучение 9, 10
Генерация волн суммарной и разностной частот 773
— гармоник 769, 773
Геометрические законы отражения и преломления 426, 543
— — на границе металла 469
Гигантские импульсы 760
Главная оптическая ось 73
Главное зеркало 185
— сечение кристалла 484
Главные нормальные сечения 102

Главные плоскости 81
— показатели преломления кристалла 524
— радиусы кривизны 102
— скорости света в кристалле 524
Главный азимут 474
— угол падения 474
Глаз близорукий 142
— дальнозоркий 142
— схематический 143
—, камера задняя 139
—, — передняя 139
—, область аккомодации 141
—, оптическая система 139
—, острота зрения 146
Глазная линза 178
Глубина проникновения 429, 470
— резкости 99
Голограмма 363
Голография 363
Гравитационное смещение спектральных линий 702
Группа волн 61
Гюйгенса принцип 24, 537
— — в формулировке Кирхгофа 304, 307
— — — — Френеля 278

Двойное лучепреломление 486
— — круговое 608
— — при механических деформациях 518
— — — течении жидкостей 519
Дебаеграмма 417
Дебая–Шерера–Хелла метод 417
Демодуляция света 209
Денисюка метод в голографии 373
Диафрагма апертурная 96
— поля зрения 98
Диоптрия 91
Дисперсионная область 330
Дисперсия аномальная 552
— временная 619
— диэлектрических осей 522
— нормальная 552
— относительная 116
— отрицательная 563
— плазмы 569
— положительная 563
— пространственная 520, 526, 527, 619
— света 40
— средняя 116
— угловая 330
— коэффициент 115
Дисторсия 112
Дисторсия, условие отсутствия 120
Дифракционная решетка 320, 325
амплитудная 325
Дифракционная решетка вогнутая отражательная 350
— — двумерная 407
Рэлея 357
— — трехмерная (пространственная) 407
— —, второстепенные максимумы 321
— —, главные максимумы 321
— —, дисперсионная область 330
— —, метод скользящего падения 326
— —, период 320, 410
— —, порядок спектра 321
— —, пропускаемость (амплитудная прозрачность) 354
— —, разрешаемое расстояние 331
— —, разрешающая способность 331, 339–349, 379
— —, — — в рентгеновской области 418, 419
Дифракция 13, 277
— рентгеновских лучей 407
— Фраунгофера 293
— — на двумерной решетке 407
— — — круглом отверстии 316
— — — отверстиях 315
— — — трехмерной решетке 407
— Френеля 293
— — от края Луны 303
— — — — экрана 302
— —, метод подобия 304
Дихроизм 495
— круговой 611
Диэлектрическая проницаемость динамическая 41
— — комплексная 468
— — металла 467
— — поперечная 483
— — продольная 483
— — статическая 41
— — эффективная 478
Диэлектрические оси 521
Диэлектрический тензор 481, 520
Длина тубуса механическая 174
— — оптическая 174
Дополнительные экраны 297
Доплера эффект 688, 690, 691, 695
Друде метод 473
Дуализм волн и частиц 32

Естественная ширина спектральных линий 577
Естественное затухание 577

Желтое пятно 140
Жесткая фокусировка 92

Закон дисперсии волн 62
— независимости световых пучков 12, 13
— обратных квадратов 158
— отражения света 12, 14
— преломления света 12, 14
— прямолинейного распространения света 12, 24
— сохранения импульса–энергии 710
— фотометрического сложения 237
Замедление хода движущихся часов 682
— — часов гравитационное 701
Зеемана эффект 596
— — простой (нормальный) 602
— — сложный (аномальный) 602
— —, классическая теория 598
— —, пи-компоненты 597
— —, сигма-компоненты 597
Зеленый луч 36
Зельмейера формула 551
Зенитное расстояние 35
Зонная пластинка 286
Зрачок глаза 97, 140
Зрение периферическое 146
— фовеальное 146
Зрительные трубы и телескопы 182

Избирательное отражение 475
— поглощение 475
Излучательная способность 717
Излучение Вавилова–Черенкова 271
— видимое 9, 10
— вынужденное (индуцированное) 743
— инфракрасное 9, 10
— переходное 276
— равновесное (черное) 713, 719
— рентгеновское 9, 10
— спонтанное 743
— температурное 717
— ультрафиолетовое 9, 10
—, спектральная плотность 230
Измерение угловых диаметров звезд 303, 402
Изображение вторичное 393
— действительное 67
— мнимое 67
— обратное 81
— оптическое 67, 103
— первичное 393
— прямое 81
— стигматическое 67
Изохроматические линии (изохроматы) 516
— поверхности 516
Иммерсия 387
Инвариантность 706
Инверсная заселенность 563
Индикатриса 545
Интенсивность излучения 716
— — удельная 716
— колебаний (света) 155, 200
Интервал 677
— времениподобный 678
— пространственноподобный 678
— световой 678
Интерференционное отражение 266
— рассеяние 453
Интерференционные полосы 204
— — равного наклона 247
— — равной толщины 244
— —, видность 223, 237
— —, локализация 245, 247
— —, полуширина 261
— —, способ наблюдения объективный 214
— —, — — субъективный 214
— —, ширина 204
— —, условие контрастности 218, 219
Интерференционный член 201
Интерференция 14, 201
— в пленках 242
— многолучевая 258
— поляризованных лучей 508, 515
—, порядок 231
—, угол 219
Интерферометр Жамена 249
— Майкельсона 256
— — звездный 404
— Маха–Цендера 256
— Рождественского 256
— Фабри–Перо 263, 339
Ионосфера 571
Искривление плоскости изображения 110
Источники света мнимые 67

Камера-обскура 13, 394
Кандела 156
Каратеодори теорема 136
Кардинальные точки 81
Картезианский овал 55
Каустика 103
Квант света 737
Керра постоянная 585
— электрооптический эффект 583
— — —, теория Борна 590

Керра электрооптический эффект, теория Ланжевена 586
Кеттелера инварианты 471
Кирхгофа закон 719
Кирхгофа–Клаузиуса формула 722
Ковариантность 706
Когерентность 201, 237
— временная 233, 238
— полная 237
— пространственная 220, 222, 238
— частичная 237
—, время 233, 238
—, длина 233, 238
—, объем 238
—, площадь 223
—, степень 236
—, — комплексная 236
Колбочки 143
Колебание квазимонохроматическое 234
— модулированное 234
Коллектив (полевая линза) 178
Коллинеарное соответствие 79
Кома 109
Компенсатор 499
Конденсор 182
Коническая рефракция 541, 543
— —, конус 538, 539
Корню спираль 299, 302
Коррекционная оправа 177
— пластинка 189
Косинусов теорема 130
Коттона–Мутона постоянная 594
— — эффект 593
Коэффициент затухания (поглощения) 582
— отражения 433
— преломления 15
— пропускания 434
— рассеяния 636
— эллиптичности 463
Кристаллическая решетка 410
Кристалл-рефрактометр 489
Кристаллы двуосные 485, 527, 536
— одноосные 483, 485, 527, 536
— отрицательные 489
— положительные 489
Критическая опалесценция 641
— температура смешения 642
Кундта постоянная 613

Лагранжа–Гельмгольца инвариант 77
— — теорема 77, 124
Лазер 748
Лазер гелий-неоновый 762
— рубиновый 754
—, трехуровневая схема 760
—, четырехуровневая схема 760
Ламберта закон 159
Ландау и Плачека формула 647
Лауэ условия 410
Лауэграмма 411
Линза магнитная 191
— одиночная 197
— с непрерывным показателем преломления 195
— толстая 93
— тонкая 77
—, фокусное расстояние 78, 198
—, формула 78, 83
Линии кривизны 102
Линия прямого зрения 141
Ллойда зеркало 213
Локальная связь 617
Лорентца–Лоренца формула 554
Лорентца преобразование 675
— сокращение длины 681
Луммера–Бродхуна фотометр 161
Луммера–Геркс пластинка 266
Лупа 172
Луч 14, 47, 528
— главный 97, 102
— мнимый 70
— необыкновенный 486
— обыкновенный 486
— параксиальный 75
Лучи остаточные 476
Лучистый поток 153
— —, интенсивность (лучистость) 153
Люкс 158
Люмен 156

Магнитная проницаемость динамическая 41
— — статическая 41
Мазер 749
Майкельсона опыт 659
— эшелон 334
Макалюзо–Корбино эффект 615
Максвелла закон 39
— эффект 519
Малюса закон 422
— теорема 56
Малых ступеней метод 148
Мандельштама Бриллюэна явление 643
— — — — вынужденное 648
Масса покоя 709

Масса релятивистская 709
Маха волны 275
Менисковые системы Максутова 188
Меслина опыт 220
Метод крюков Рождественского 566
Механика релятивистская 706
Механический эквивалент света 157
Микроволны 9
Микроскоп 174, 182, 387
Мировая линия 677
— точка 677
Многослойные покрытия 267
Модуляция амплитудная 234
— фазовая 234
Молекулы полностью анизотропные 586
— полярные 556
Монохроматическое векторное поле 423
Монохроматы 148
Мосотти–Клаузиуса формула 554
Мутная среда 632

Направления синхронизма 772
Некогерентность 201, 237
Нелокальность временная 617
— пространственная 617
Ньютона кольца 247

Обратная связь 751
Объектив иммерсионный 125
Объекты абсорбционные 399
— рефракционные 400
Одновременность 667
Окуляр Гюйгенса 178
— Кельнера 180
— компенсационный 176
— Рамсдена 178, 180
Оптика 9
— геометрическая 12, 44
— ионная 191
— линейная 42
— нелинейная 42, 748
— параксиальная (гауссова) 78
— физическая 9
— электронная 69, 191
Оптическая накачка 755
— область спектра 10
— сила 91
Оптические оси 483
— — второго рода (бинормали) 527
— — первого рода (бирадиали) 535
Оптический интервал 89
Оптическое детектирование 768
Ортотомная система 56
Освещенность 157
Освещенность естественная 166
Освещенность энергетическая (облученность) 157
Относительное отверстие 165
Отражение диффузное 17
избирательное 267
интерференционное 266
— правильное 16
Очки 143

Палочки 143
Парадокс близнецов 687
Параметрическая генерация света 776
Параметрическое усиление света 777
Перспектива 100
Плазменная (ленгмюровская) частота 570
Планка постоянная 10, 739
— формула 739
Пластинка в четверть волны 499
— — — —, главные направления 499
Плеохроизм 495
Плоскость поляризации (колебаний) 38, 420, 433
— —, вращение естественное 605, 620
— —, — магнитное (эффект Фарадея) 605, 611
— установки (наводки) 99
Плотность лучистой энергии 153, 715
— — — спектральная 715
Поверхностная прозрачность 434
Поверхность волновых векторов 545
— лучевая (волновая) 529
— нормалей 525
— показателей преломления 545
— равных амплитуд 43
— — фаз 43
Поглощательная способность 717
Показатель затухания 470
— — главный 468
— преломления 470
— — главный 468
— — комплексный 468
— — межпланетного и межзвездного пространств 453
— — необыкновенный 486
— — обыкновенный 486
— — рентгеновских лучей 453
Поккельса линейный электрооптический эффект 594
Покровное стекло 177
Полное отражение 16, 429, 436, 546
— —, предельный угол 16, 429

Полуширина спектральной линии 578
Поля опыт 214
Поляризатор 421
Поляризация вакуума 30
Поляризация света круговая 423
— — линейная 39, 420
— — эллиптическая 422, 473
— — — при полном отражении 438
Поляризуемость 766
Поляроиды 421, 495
Правило сумм 562
Прево правило 719
Призма дисперсионная 119
— — прямого зрения 119
— — сложная 119
— — разрешающая способность 339, 349
— поляризационная Аренса 493
— — Волластона 495
— — Глазебрука 493
— — Глана 493
— — Глана–Томпсона 493
— — Николя (николь) 491
— — Рошона 495
— — Сенармона 495
— — Франка–Риттера 493
— — Фуко 493
Призмы поляризационные двухлучевые 491
— — однолучевые 491
Принцип детального равновесия 713
— неопределенности 229
— суперпозиции 199
— таутохронизма 70
Просветление оптики 444
Просветляющие фильтры 759
Пространство изображений 69
— предметов 69
Пуркинье явление 149
Пучок лучей астигматический 103
— — гомоцентрический 67
— — опорный 364
— — предметный 364

Радиоволны 9
Радужная оболочка 139
Разность хода 205
— — оптическая 205
Разрешаемое расстояние 146, 379, 383, 386, 391, 393
Разрешающая способность 146
— — глаза 384
— — голограммы 393
— — микроскопа 386, 391, 393
— — объектива 378
Разрешающая способность спектральная 230, 263, 348
— — телескопа 383
— —, критерий Рэлея 332
Рассеяние света 13, 277, 632
— — молекулярное (рэлеевское) 637
— — —, деполяризация 639
— — —, тонкая структура 641
— — — комбинационное 650
— — — — вынужденное 653
— — — поверхностное 642
— — тиндалевское 632
Расстояние ясного зрения 142
Рефлектор 185
Рефрактометр интерференционный 251
Рефрактор 185
Рефракция 34
Рефракция атомная 555
— молекулярная 555
— удельная 554
Роговица 139
Родопсин 151
Рождение пар 30
Роулэнда круг 352
«Рыбий глаз» 139
Рэлея–Джинса формула 735
Рэлея закон 636
— формула 639

Самоканализация 775
Самофокусировка 775
Сателлиты 650
— красные (стоксовы) 651
— фиолетовые (антистоксовы) 651
Свет естественный 420
— квазимонохроматический 208, 234
— поляризованный 420
— — частично 420
Светимость 160
— энергетическая (излучаемость) 160
Световая эффективность 156
Световое давление 558
— поле вблизи фокуса 374
Световой вектор 269
— конус 680
— поток 155
Световые (фотометрические) величины 155
Светосила 165
Свеча международная 156
Связь массы с энергией 709
Сетчатка (ретина) 140, 143

Сила осциллятора 561
— света источника 155
— — — средняя сферическая 156
Система афокальная (телескопическая) 81, 84
— диоптрическая 83
— катоптрическая 83
Скачок фазы при отражении 234, 243, 432
— — — — полном 438
Скин-эффект аномальный 479
Скин-эффект нормальный 478
Склера 139
Скорость групповая 57, 60, 63, 528
— движения энергии 64
— лучевая 528
— нормальная 55, 528
— передового фронта 65
— света в вакууме 667
— фазовая 58
Слепое пятно 144
Сложение скоростей релятивистское 703
Собственная длина 681
Собственное время 683
События абсолютно будущие 679
— — прошедшие 679
— — удаленные 679
Сопряженные точки 67
Сопутствующая система отсчета 683
Сопутствующие часы 683
Сосудистая оболочка 139
Спектр нормальный 330
— призматический 330
Статистически независимые величины 626
Стационарные состояния 89
Стекловидное тело 139
Степень поляризации 446
Стереографическая проекция 137
Стереотруба 184
Стефана–Больцмана закон 724
Стефана–Больцмана постоянная 742
Стильб 159
Столетова стопа 433
Сферическая аберрация 107
— — поперечная 108
— — продольная 108

Тахионы 680
Телескоп Гершеля–Ломоносова 186
— Грегори 187
— Кассегрена 187
— Ньютона 186
Телефон световой 437
Тени 13
Теорема Ван-Циттера–Цернике 242
— обращения 532
Теория света волновая 23
Теория света корпускулярная 21
— — электромагнитная 29
Ток поляризации 467
— проводимости 467
Точка ясного видения ближняя 141
— — — дальняя 141
Точки апланатические 122
— главные 81
— — обратные 85
— закругления 102
— ортоскопические 121
— узловые 84
— — обратные 85
— фокальные 80, 102
Третичный спектр 118
Труба Галилея 184
— Кеплера 183
Турмалин 495

Увеличение зрительной трубы 87
— лупы 173, 388
Увеличение микроскопа 170, 172, 388
— нормальное 167
— осевое (продольное) 85
— поперечное 81
— угловое 84
Угол аберрации 693
— зрения 141
— преломления вещественный 470
— проекции 97
— раскрытия 97
— рассеяния 634
— синхронизма 772
Уголковый отражатель 18
Ультрамикроскоп 387
Ультрафиолетовая катастрофа 736
Уравнение адиабаты для излучения 725
Ускоритель на встречных пучках 711
Условие Гершеля 131
— синусов 122, 131
— тангенсов (ортоскопии) 121
Уширение спектральных линий доплеровское 581
— — — естественное 577
— — — ударное 580
— — — штарковское 581

Фазовая пластинка 402
Фазовый контраст 399
— — негативный 402

Фазовый контраст позитивный 402
— синхронизм 771, 777, 778
Ферма принцип 49
Флуктуация 626
— относительная 626
— среднеквадратичная 626
Фокальные отрезки 103
— плоскости 80
Фокус 67
— главный 80
— параксиальный 106
Формула зеркала 75, 83
Фотон 32
Френеля бипризма 212
— закон лучевых скоростей света в кристалле 533
— — нормальных скоростей света в кристалле 523
— зеркала 211
— коэффициенты 431
— параллелепипед 439
— формулы 431, 472
Функция автокорреляционная 236
— корреляционная 236
— — взаимная 236
Фурье интеграл 226
Фурье ряд 225

Характеристическая функция 129
Хроматическая поляризация 513
Хрусталик 139, 141

Цветная фотография 270
— Липпмана 269
Цветного зрения теория 147
Центр интерференционной картины 230
Центральная ямка 141
Центрированная оптическая система 73

Часы идеальные 683
Черенковские счетчики 275
Четырехмерный вектор 707

Шмидта камера 188
Штарка эффект 602
Шустера зоны 298

Эйконал 46
точечный 129
, уравнение 46
Эйнштейна коэффициенты 744
— принцип относительности 658
— формула 641
Эйри кружок 378
— спирали 518
Электроны оптические 548
Энантиоморфизм 606
Энергетическая сила источника 155
Энергетические величины 155
Энергия кинетическая релятивистская 709
покоя 709
— полная 709
Эрмитов тензор 520
Эфир световой (мировой) 23
Эффективная температура излучения 762

Юнга опыт 210

Яркость 158
— энергетическая 158

Учебное издание

СИВУХИН Дмитрий Васильевич

ОБЩИЙ КУРС ФИЗИКИ

Том IV
ОПТИКА

Редактор *Д. А. Миртова*
Оригинал-макет: *О. Б. Широкова*
Оформление переплета: *А. А. Логунов*

ЛР № 071930 от 06.07.99
Подписано в печать 14.01.05. Формат 60×90/16
Бумага офсетная № 1. Печать офсетная
Усл. печ. л. 49,5. Уч.-изд. л. 51,17. Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117864 Москва, Профсоюзная, 90
E-mail: fizmat@maik.ru; http://www.fml.ru

Отпечатано с диапозитивов
в РГУП «Чебоксарская типография № 1»
428019 Чебоксары, пр. И. Яковлева, 15

ISBN 5-9221-0228-1

9 785922 102285